Gregory Kordas
Last update: March 22, 2023

LECTURE 3
UNIVARIATE PROBABILITY DISTRIBUTIONS

Abstract: We introduce the concept of a random variable, discuss discrete and continuous variables, their transformations and convolutions, as well as, their moments, moment generating functions and characteristic functions.

1. Random variables

The essential reason why random variables and associated quantities are introduced springs from the fact that the underlying sample spaces on which σ-fields and probability measures are defined are arbitrary. A *probability measure* P is a set function that assigns probabilities to events like a side of a die or the face of a coin, and so on. However, further progress needs to be made. For example, it is not possible to add a head on a coin to a red card chosen from a pack of cards. A second consideration is that there is a kind of equivalence between many experiments. For example, we may well assign similar probabilities to the outcome of a head on the toss of a coin, the occurrence of an odd number on rolling a die and the drawing of a red card from a pack, yet in terms of the underlying Ω these experiments are obviously different. As a result of considerations like these, it is often desirable to cast the sample space Ω itself into the background, and contact our analysis in terms of other quantities that are easier to handle mathematically and generalize directly to other experiments with similar structure.

The first step in liberating our analysis from the limitations of the sample space Ω is to define a random variable on this space.

Definition: Let $\mathscr{F}$ be a σ-field defined on the sample space Ω, and let $\mathscr{B}$ be the Borel σ-field on the real line $\mathbb{R}$. A *random variable* $X(\cdot)$ is a function $X(\omega) : \Omega \mapsto \mathbb{R}$ mapping the sample space into the real line, such that for all $B \in \mathscr{B}$, $X^{-1}(B) \in \mathscr{F}$.

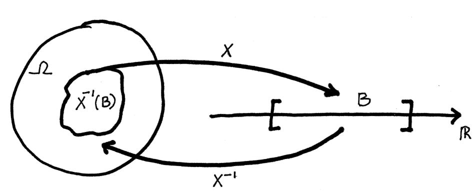


Figure 1. The measurability of the random variable X.

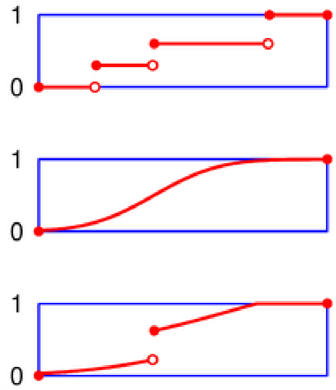


FIGURE 2. Types of cdf's: (a) discrete, (b) absolutely continuous, (c) mixed.

The last condition in the definition of a random variable, i.e. that for all $B \in \mathscr{B}$, $X^{-1}(B) \in \mathscr{F}$, means that $X(\cdot)$ is a *measurable* function from Ω to $\mathbb{R}$. Measurability is defined as requiring that the inverse image of X is an element of the σ-field $\mathscr{F}$, i.e., an event.

A *cumulative distribution function* (c.d.f.) is a real valued function

$$F(x) = P(X \leq x),$$

defined on the real line, having the following properties:

(i) $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$,
(ii) $F(+\infty) = \lim_{x \to +\infty} F(x) = 1$,
(iii) $\lim_{\epsilon \downarrow 0} F(x + \epsilon) = F(x)$,
(iv) If $x_1 < x_2$ then $F(x_1) \leq F(x_2)$.

That is, every c.d.f. increases monotonically from zero to one and is right continuous We say that $F(x)$ is a *cadlag* (French for "continue á droite, limite á gauche") function, i.e., it is continuous on the right with limits on the left. For each $\epsilon > 0$, a cadlag function has, at most, finitely many discontinuities of magnitude greater than $\epsilon > 0$ in any compact interval. Otherwise, these discontinuities would have a right or left limit point, destroying the cadlag property. Since the real line can be covered by a countable number of compact sets, a cadlag function can have, at most, a countable number of discontinuities. Conversely, every real valued function $F(x)$ on the real line which has the properties specified above is the distribution function of some random variable. The distribution functions are classified into four main types: (a) discrete, (b) absolutely continuous, (c) mixed, and (d) singular.

2. Discrete Random variables

We start by considering random experiments that admit only a countable (finite or infinite) number of elementary outcomes ω, and define a random variable X on these outcomes, with values $-\infty < \xi_1, \xi_2, ... < \infty$. Then, $X(\omega)$ is a *discrete* random variable, and its possible values $\xi_1, \xi_2, ...$ are its *mass points.* Probabilities $P\{X(\omega) = x\}$ are assigned according to a real valued function $f(\cdot)$, defined on the real line and called the *probability mass function* (p.m.f.), with the following properties:

(i) $f(x) \geq 0$, for all x;
(ii) $f(x) = 0$ except at the mass points $\xi_1, \xi_2, ...$;
(iii) $\sum_i f(\xi_i) = 1$.

Conversely, every real valued function $f(x)$ on the real line which has these properties is the pmf of some random variable.

The following discrete distributions are of particular interest.

(1) *Bernoulli* with parameter p, $(0 \leq p \leq 1)$.

$$f(x) = p^x(1-p)^{1-x}, \quad \text{for } x = 0, 1.$$

This random variable describes the toss of a coin, in which case $X = 1$ if Heads, $X = 0$ if Tails, the probability of Heads is $\Pr(X = 1) = p^1(1-p)^0 = p$, and the probability of Tails is $\Pr(X = 0) = p^0(1-p)^{1-0} = (1-p)$. Alternatively, X could be the employment status of an individual, with p being the probability of employment and $(1-p)$ being the probability of unemployment.

(2) *Binomial* with parameters n, p (n positive integer, $0 \leq p \leq 1$).

$$f(x) = \frac{n!}{x!(n-x)!} p^x(1-p)^{n-x}, \quad \text{for } x = 0, 1, 2, ..., n.$$

This variable describes the toss of n identical coins. For instance, letting $X =$"# of Heads", for $n = 1$ the Binomial reduces to the Bernoulli, while for $n = 2$, we compute, $f(0) = (1-p)^2$, $f(1) = 2p(1-p)$, and $f(2) = p^2$.

(3) *Discrete Uniform* with parameter n (a positive integer).

$$f(x) = \frac{1}{n}, \quad \text{for } x = 1, 2, ..., n.$$

(4) *Poisson* with parameter $\lambda > 0$.

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad \text{for } x = 0, 1, 2, \ldots$$

The Poisson random variable describes the number of flights arriving at an airport in the space of 24 hours, or the number of patent applications submitted in a year. The Taylor series expansion of the exponential e^{λ} is given by

$$e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = 1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6} + \cdots,$$

so that $\displaystyle\sum_{x=0}^{\infty} f(x) = 1$, as required.

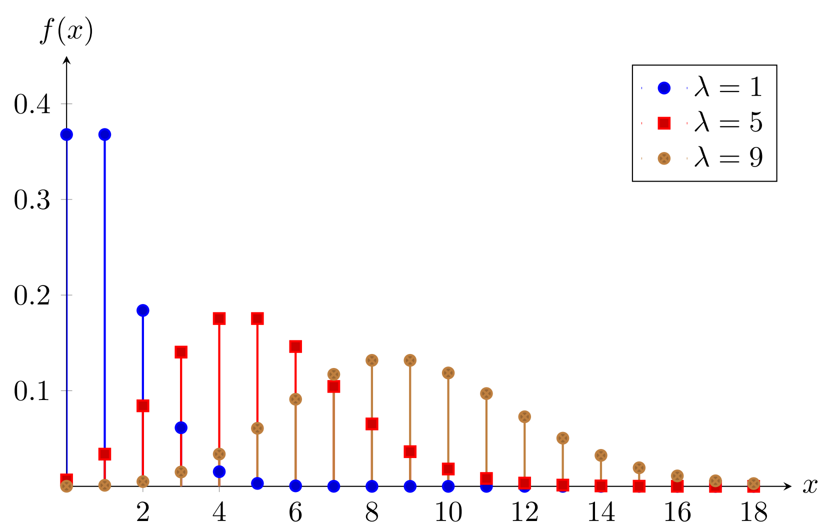


FIGURE 3. Poisson p.m.f. for $\lambda = 1, 5, 9$.

3. ABSOLUTELY CONTINUOUS RANDOM VARIABLES

To assign probabilities we use the *probability density function* (p.d.f.) $f : \mathbb{R} \to \mathbb{R}$, continuous, with the following properties:

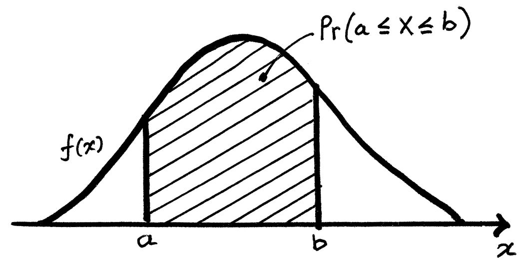
(i) $f(x) \geq 0$ everywhere;

(ii) $\displaystyle\int_{-\infty}^{\infty} f(x)\, dx = 1$.

The value $f(x)$ assigned to any particular x, is not a probability *per se*, since in the continuous case the probability of any given value x is necessarily zero. Instead, the pdf is used to compute probabilities as areas under the curve,

$$\Pr(a \leq X \leq b) = \int_a^b f(x)\,dx.$$

That is, the probability that the random variable X will take values in the interval $[a, b]$ is given by the area under f between a and b. Since f is a continuous function, $\Pr(a < x < b) = \Pr(a \leq X \leq b)$ and thus $\Pr(X = a) = \Pr(X = b) = 0$, as required.



Of particular interest in applications is the probability of the event $\{-\infty < X \leq a\}$. We define the *cumulative distribution function* (cdf), $F(x)$, as

$$F(x) = \Pr(-\infty < X \leq x) = \int_{-\infty}^{x} f(t)dt.$$

The cdf gives the area under the pdf from $-\infty$ to x. Some properties of the cdf follow immediately from the definition:

(i) $F(-\infty) = 0, F(\infty) = 1$;
(ii) $F(x)$ is nondecreasing (since $f(x) \geq 0$);
(iii) Whenever differentiable, $dF(x)/dx = f(x)$, because $F = \int f(t)dt$, and the derivative of an interval with respect to its upper limit is just the integrand evaluated at the upper limit.

The following continuous distributions occur often in applications:

(1) *Uniform* on the interval $[a, b]$ $(a < b)$.

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{for } a \leq x \leq b \\ 0, & \text{otherwise.} \end{cases}$$

Observe that $f(x) \geq 0$ everywhere, and

$$\int_{-\infty}^{\infty} f(x)dx = \int_a^b \frac{1}{b-a}dx = \frac{1}{b-a}x\bigg|_a^b = 1,$$

as required. Also,

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0, & \text{for } x \leq a \\ \dfrac{x-a}{b-a}, & \text{for } a \leq x \leq b \\ 1, & \text{for } x > b. \end{cases}$$

(2) *Exponential* with parameter $\lambda > 0$.

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{for } x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

We compute

$$\lambda \int e^{-\lambda t}dt = \lambda \frac{e^{\lambda t}}{-\lambda} = -e^{\lambda t}$$

so,

$$F(x) = \begin{cases} 0, & \text{for } x \leq 0, \\ 1 - e^{-\lambda x}, & \text{for } x > 0. \end{cases}$$

The exponential distribution may be appropriate for modelling the life of many electronic devices, or for the duration of unemployment spells.

(3) *Standard normal.* The pdf, denoted by $\phi(x)$, is given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \quad x \in \mathbb{R}.$$

Clearly, $\phi(x) > 0$ for all $x \in \mathbb{R}$. To see that it also integrates to one, consider

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-x^2/2}dx,$$

and let $x^2/2 = z$ and $dx = z^{-1/2}/\sqrt{2}\,dz$ to obtain

$$I = \frac{1}{\sqrt{\pi}} \int_0^{\infty} z^{-1/2}e^{-z}dz.$$

The integral in the rhs is equal to $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ (see the discussion of the Gamma function below), so that $I = 1$ and $\phi(x)$ is a proper density.

The standard normal p.d.f. $\phi(x)$ has the following properties:

(i) $\phi(0) = 1/\sqrt{2\pi} \approx 0.3989$;
(ii) $\phi(x) = \phi(-x)$;
(iii) $\phi'(x) = -x\phi(x)$;

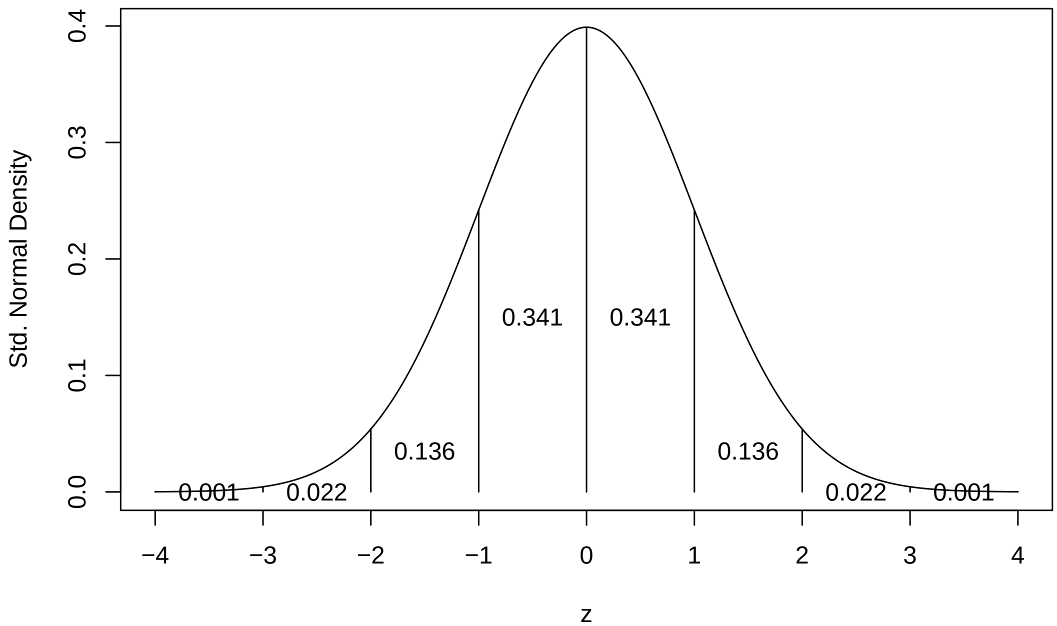


FIGURE 4. The standard normal density. In a normal population approximately 68% of the observations are within one s.d. from the mean, and 95% are within 2 s.d.'s from the mean. The probability beyond ± 3 s.d.'s is almost zero.

(iv) $\phi''(x) = (x^2 - 1)\phi(x)$, so $\phi(x)$ has inflection points at $x = \pm 1$.

The c.d.f. of the standard normal

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt, \qquad -\infty < x < \infty.$$

is not available in closed form (it cannot be written in terms of elementary functions), but it is widely tabulated. Note that tabulations of $\Phi(x)$ for $x > 0$ are sufficient, since by symmetry $\Phi(-x) = 1 - \Phi(x)$. By a series expansion of $e^{-t^2/2}$ and direct integration, one can immediately obtain the formula

$$\Phi(x) = \frac{1}{2} + \frac{1}{2\pi} \sum_{j=0}^{\infty} \frac{(-1)^j x^{2j+1}}{2^j (2j+1)}, \qquad -\infty < x < \infty.$$

Computations, however, using this formula are often inefficient. An excellent approximating formula due to Zelen and Severo (1968)[1] is

$$\Phi(x) = 1 - \phi(x)\left[b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5\right] + \epsilon(x), \quad x \geq 0,$$

where $t = (1 + px)^{-1}$, $p = .2316419$; $b_1 = .319381530$; $b_2 = -.356563782$; $b_3 = 1.781477937$; $b_4 = -1.821255978$; $b_5 = 1.330274429$. The magnitude of the approximation error is $|\epsilon(x)| < 7.5 \times 10^{-8}$, i.e., whithin computer single-precision.

(4) *Standard logistic.* The p.d.f. is given by

$$f(x) = \frac{e^x}{(1 + e^x)^2}, \qquad -\infty < x < \infty,$$

and the c.d.f. is given by

$$F(x) = \int_{-\infty}^{x} f(t)dt = \frac{e^x}{1 + e^x}, \qquad -\infty < x < \infty.$$

The logistic pdf is very similar in shape to the normal pdf (both are bell-shaped, unimodal and symmetric about their mean), but the the logistic density has slightly thicker tails than the normal. An interesting property of the logistic distribution is that it is the unique solution of the following differential equation,

$$f(x) = \frac{dF(x)}{dx} = F(x)(1 - F(x)).$$

(5) *Standard Cauchy.* The pdf is given by

$$f(x) = \frac{1}{\pi(1 + x^2)}, \quad -\infty < x < \infty.$$

Cauchy random variables are notoriously badly behaved, and for this reason they play an important theoretical role as counterexamples to many results that hold for well behaved variables but fail to hold for Cauchy variables.

Two important families of distributions are the *Beta* and *Gamma families.* Associated with them are the Beta and Gamma functions. The *Gamma function* is defined as

$$\Gamma(t) = \int_0^\infty u^{t-1} e^{-u} du, \quad \text{for } t > 0.$$

[1] Zelen, M. and Severo, N. C. (1964), "Probability Functions", Ch. 26. in *Handbook of Mathematical Functions*, M. Abramowitz and I. A. Stegun (eds), 925-995, U. S. Department of Commerce, Applied Mathematics Series.

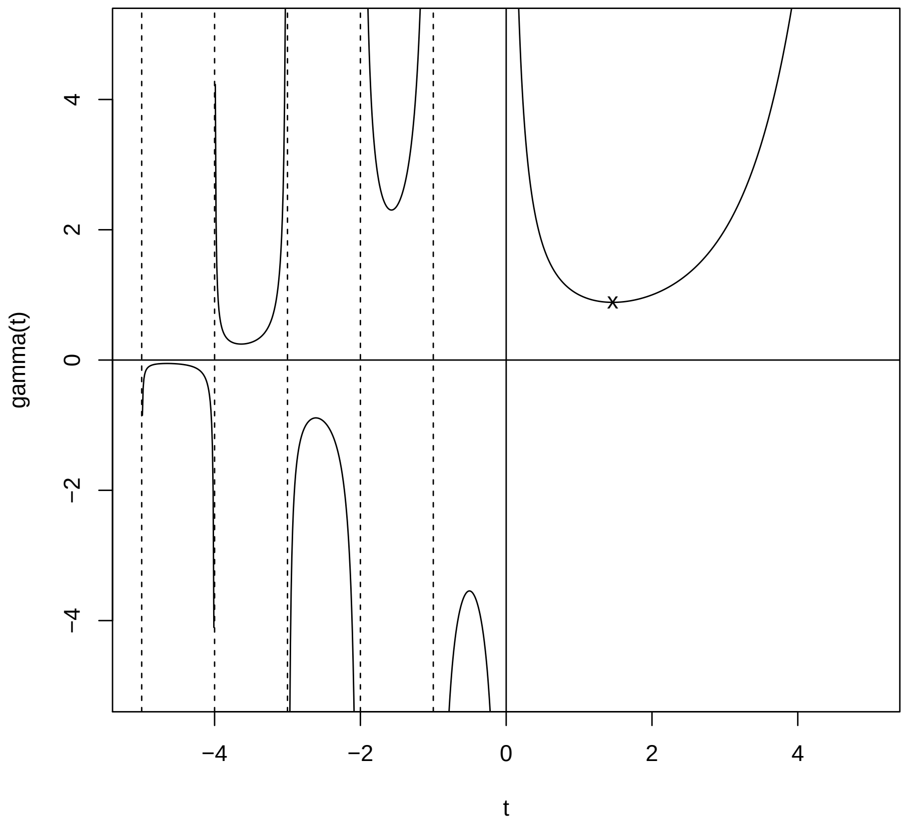


FIGURE 5. The $\Gamma(t)$ function. The minimum for $t > 0$ occurs at $t = 1.46163...$ marked by $\times$.

THEOREM 1. *The following properties of the Gamma function are used frequently,*
(i) $\Gamma(1) = \Gamma(2) = 1$
(ii) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ (isn't this amazing?!)
(iii) $\Gamma(t) = (t-1)\Gamma(t-1)$ for $t > 1$ and real
(iv) $\Gamma(t) = (t-1)!$ for t positive integer.

Proof: Part (iii) follows from integrating $\Gamma(t)$ by parts:

$$\Gamma(t) = -u^{(t-1)}e^{-u}]_{0+}^{\infty} + (t-1)\int_{0+}^{\infty} u^{t-2}e^{-u}du.$$

The first term in the r.h.s. is zero, and the second is $(t-1)\Gamma(t-1)$, $t>0$. For example,

$$\begin{aligned}\Gamma(\tfrac{7}{2}) &= (\tfrac{5}{2})\Gamma(\tfrac{5}{2})\\ &= (\tfrac{5}{2})(\tfrac{3}{2})\Gamma(\tfrac{3}{2})\\ &= (\tfrac{5}{2})(\tfrac{3}{2})(\tfrac{1}{2})\Gamma(\tfrac{1}{2}),\end{aligned}$$

a sort of generalized factorial. Clearly, for t positive integer (iv) holds. (i) is a special definition that holds for factorials too: $0! = 1! = \Gamma(2) = (2-1)! = 1$. Part (ii) is of special importance and the only difficult part to prove. We have

$$\Gamma(\tfrac{1}{2}) = \int_0^\infty u^{-\frac{1}{2}} e^{-u} du.$$

Letting $u = z^2$ so that $du = 2z\,dz$ the integral takes the alternate form

$$\Gamma(\tfrac{1}{2}) = 2\int_0^\infty e^{-z^2} dz.$$

This is the so-called *probability integral*, called by this name because it is related to the area under the normal curve. To evaluate it, let

$$I = \int_0^\infty e^{-z^2} dz = \int_0^\infty e^{-y^2} dy,$$

and revolve the curve $x = e^{-y^2}$ about the X-axis. The surface generated by the revolution has the equation $x = e^{-(y^2+z^2)}$. The volume between this surface and the YZ-plane is

$$\begin{aligned}V &= 4\int_0^\infty \int_0^\infty e^{-(y^2+z^2)} dydz\\ &= 4\int_0^\infty \left[\int_0^\infty e^{-y^2} dy\right] e^{-z^2} dz\\ &= 4\int_0^\infty A\, e^{-z^2} dz = 4A^2.\end{aligned}$$

By the *method of hollow cylinders*, this volume can be expressed as

$$V = 2\pi \int_0^\infty y e^{-y^2} dy = \pi.$$

Equating the two results, we get $4I^2 = \pi$, or $I = \frac{1}{2}\sqrt{\pi}$, and (ii) is proved.

Second proof[2]: We wish to show that the Gaussian integral

$$I = \int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}.$$

[2]Georgakis G. (1994), "A Note on the Gaussian Integral", *Mathematics Magazine*, vol. 67, p. 47.

Let $y = xs$, $dy = xds$, then

$$\begin{aligned} I^2 &= \int_0^\infty \left(\int_0^\infty e^{-(x^2+y^2)} dy \right) dx = \int_0^\infty \left(\int_0^\infty e^{-x^2(1+s^2)} x\, ds \right) dx \\ &= \int_0^\infty \left(\int_0^\infty e^{-x^2(1+s^2)} x\, dx \right) ds \\ &= \int_0^\infty \left[\frac{1}{-2(1+s^2)} e^{-x^2(1+s^2)} \Big|_0^\infty \right] ds = \frac{1}{2} \int_0^\infty \frac{ds}{1+s^2} \\ &= \frac{1}{2} \arctan s \Big|_0^\infty = \frac{\pi}{4}. \end{aligned}$$

■

Related to the probability integral discussed in the proof above, is the so-called *error function* $\mathrm{erf}(x)$ defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \qquad -\infty < x < \infty.$$

This function is related to the normal c.d.f. $\Phi(x)$ by

$$\Phi(x) = \frac{1}{2} + \frac{1}{2} \mathrm{erf}(x/\sqrt{2}).$$

The *Beta function* is defined by

$$B(p,q) = \int_0^1 u^{p-1}(1-u)^{q-1} du, \quad \text{for } p > 0, q > 0,$$

and it is not difficult to prove that

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

(6) *Gamma distribution* with parameters $p > 0, \lambda > 0$. The p.d.f. of a $\gamma(p, \lambda)$ is given by

$$f(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x}, \quad x > 0.$$

The parameter p controls the shape of the distribution, while λ controls its scale. It can be shown that if $X_1 \sim \gamma(p_1, \lambda)$ and $X_2 \sim \gamma(p_2, \lambda)$ independent of X_1, then $X_1 + X_2 \sim \gamma(p_1 + p_2, \lambda)$. Also, $\gamma(\nu/2, \nu/2) = \chi^2_\nu$, i.e., the chi-square distribution with ν degrees of freedom.

(7) *Beta distribution* with parameters $p > 0, q > 0$. The pdf of a $\beta(p, q)$ is given by

$$f(x) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}, \quad x > 0,$$

It can be shown that if $X_1 \sim \gamma(p_1, \lambda)$ and $X_2 \sim \gamma(p_2, \lambda)$, independent of X_1, then

$$\frac{X_1}{X_1 + X_2} \sim \beta(p_1, p_2).$$

4. Transformations

Suppose that X has a known pdf $f(x)$, and we wish to find the pdf $g(y)$ of $Y = h(X)$. One way to approach the problem would be to first compute the cdf of Y by

$$G(y) = \Pr(Y \leq y) = \Pr(h(X) \leq y),$$

and then recover $g(y)$ by differentiating $G(y)$.

Example 1. Let

$$f(x) = \begin{cases} \frac{3}{2}x^2, & \text{for } -1 \leq x \leq 1, \\ 0, & \text{otherwise,} \end{cases}$$

and let $Y = X^2$. Then

$$\begin{aligned} G(y) = \Pr(Y \leq y) &= \Pr(X^2 \leq y) = \Pr(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} f(x)dx = \int_{-\sqrt{u}}^{\sqrt{y}} \frac{3}{2}x^2 dx \\ &= \frac{3}{2} \cdot \frac{x^3}{3}\bigg|_{-\sqrt{y}}^{\sqrt{y}} = y^{3/2}. \end{aligned}$$

Therefore,

$$g(y) = \frac{dG(y)}{dy} = \frac{3}{2}\sqrt{y}.$$

Now the range of Y can be obtained from the relationship $Y = X^2$. Since $-1 \leq x \leq 1$, we should have $0 \leq y \leq 1$. Therefore,

$$g(y) = \begin{cases} \frac{3}{2}\sqrt{y}, & \text{for } 0 \leq y \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to check that

$$\int_0^1 g(y)dy = \frac{3}{2}\int_0^1 \sqrt{y}dy = \frac{3}{2} \cdot \frac{x^{3/2}}{3/2}\bigg|_0^1 = 1.$$

■

If the function $Y = h(X)$ is one-to-one and differentiable, then the inverse function $X = h^{-1}(Y)$ exists, so we can write

$$\begin{aligned} G(y) &= \Pr(h(X) \leq y) = \Pr(X \leq h^{-1}(y)) \\ &= \int f(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right| dy, \end{aligned}$$

where $J = \dfrac{dh^{-1}(y)}{dy}$ is the *Jacobian* of the transformation and $|\cdot|$ is the absolute value function. Thus, the density of Y is given by,

$$g(y) = f(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right|.$$

Recall that the probability of X lying in an infinitesimally small region with volume dx is given by $f_x(x)dx$. Since $h(\cdot)$ can expand or contract space, the infinitesimal volume surrounding x in $X-$space may have different volume in Y-space. We thus need to preserve the property

$$|f_y(h(x))dy| = |f_x(x)dx|,$$

from which we obtain

$$f_y(y) = f_x(x) \left| \frac{dx}{dy} \right|,$$

which is the same as the formula above for $x = h^{-1}(y)$.

Example 2. (*The Normal location-scale family*). Let X be a standard normal random variable, i.e.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty$$

and let $Y = \mu + \sigma X$, for $\mu, \sigma \in \mathbb{R}$. Clearly this is an one-to-one and differentiable transformation, so the inverse function exists. We have

$$X = h^{-1}(Y) = \frac{Y - \mu}{\sigma},$$

and

$$|J| = \left| \frac{dh(y)}{dy} \right| = \frac{1}{|\sigma|}.$$

Therefore,

$$\begin{aligned} g(y) &= f(h^{-1}(y)) \left| \frac{dh(y)}{dy} \right| \\ &= f\Big(\frac{y-\mu}{\sigma}\Big) \cdot \frac{1}{|\sigma|} \\ &= \frac{1}{\sqrt{2\pi}|\sigma|} e^{-(y-\mu)^2/(2\sigma^2)}, \qquad -\infty < y < \infty. \end{aligned}$$

Therefore, if X is standard normal, $Y = \mu + \sigma X$ has the normal density with mean μ and variance σ^2. Note that it is natural to restrict σ in the positive orthant, but in any case only $|\sigma|$ is relevant here. ■

The last example demonstrates a general rule that follows directly from the transformation formula: If X has density $f(x)$, then the density of $Y = \mu + \sigma X$, for $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$, is given by

$$g(y) = \frac{1}{\sigma} f\Big(\frac{y-\mu}{\sigma}\Big).$$

Thus, given a standard density (like the standard normal, the standard logistic, or the standard Cauchy) it is easy to create an entire family of densities, often referred to as a *location-scale family*, by the above method. Note that neither μ is necessarily the mean of Y, nor σ^2 the variance of Y. The parameters μ and σ specify the general location and scale of the distribution in the sense that as they increase the mean and the variance of Y increase. For example, in the gerenalized logistic ditrubution $E(Y) = \mu$ because the distribution is symetric, but $V(Y) = \pi^2\sigma^2/3$.

Another very useful transformation is the so-called *probability transformation*: If X is a continuous random variable with cdf F, then $F(X)$ is a uniform random variable on $[0,1]$, i.e.,

$$X \sim F \Leftrightarrow F(X) \sim U[0,1].$$

Example 3. (*The probability transformation*). Let X be a standard logistic random variable with density

$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2},$$

and let

$$h(x) = F(x) \equiv \frac{e^x}{1+e^x}.$$

Then

$$x = h^{-1}(y) = \log\left(\frac{y}{1-y}\right) = \log y - \log(1-y),$$

and

$$\frac{dh^{-1}(y)}{dy} = \frac{1}{y} + \frac{1}{1-y} = \frac{1}{y(1-y)},$$

which is always positive. Now $0 \leq y \leq 1$, and

$$\begin{aligned} g(y) &= f(h^{-1}(y)) \cdot \left| \frac{dh^{-1}(y)}{dy} \right| \\ &= f\Big(\log \Big(\frac{y}{1-y} \Big) \Big) \cdot \frac{1}{y(1-y)} \\ &= \frac{e^{-\log(\frac{y}{1-y})}}{\Big(1 + e^{-\log(\frac{y}{1-y})}\Big)^2} \cdot \frac{1}{y(1-y)} \\ &= 1. \end{aligned}$$

Thus, $Y = F(X)$ is indeed distributed as $U[0,1]$. ■

One application of the probability transformation is in drawing random samples from a required distribution: if a random sample from a $U \sim U[0,1]$ is available, then $X = F^{-1}(U) \sim F$ is a random sample from F. Most computer programs include a uniform random number generator, so if F^{-1} is available, it is easy to draw samples from the distribution F. For example, in STATA the following commands will produce a sample from the logistic distribution

set obs 100	Set the sample size.
gen u = uniform()	Draw a uniform sample.
gen x = log(u/(1-u))	Generate the logistic sample.

5. EXPECTATIONS

Let X be a random variable defined on a probability space (Ω, A, P) with cdf $F(x)$. The *expectation* of X exists if and only if

$$E(|X|) = \int_{-\infty}^{\infty} |x| dF(x) < \infty \quad \text{(integrability)}$$

i.e., if X is an integrable function. The expectation of X is then defined by

$$E(X) = \int_{-\infty}^{\infty} x dF(x).$$

To understand the integrability condition, note that by the additivity of integrals, the expectation of X can be written as

$$E(X) = E(X^+) - E(X^-),$$

where

$$X^+ = X\,I\{X \geq 0\} = \max\{X, 0\}$$

is the *positive part* and

$$X^- = -X\,I\{X < 0\} = \max\{-X, 0\}$$

is the *negative part* of X. If $E(X^+) = \infty$ and $E(X^-) = -\infty$, we get $E(X) = \infty - \infty$, which is not defined! To avoid this anomaly, we restrict the expectation of $|X|$ to be finite.

Provided that at least one of $E(X^+)$ and $E(X^-)$ is finite, we can define $E(X) = E(X^+) - E(X^-)$; otherwise (that is, if both $E(X^+)$ and $E(X^-)$ are infinite), $E(X)$ is undefined (does not exist).

Definition 1. *Suppose that X is a nonnegative random variable with distribution function F. The expected value or mean of X (denoted by $E(X)$) is defined to be*

$$E(X) = \int_0^\infty [1 - F(x)]dx$$

which may be infinite. In general, if $X = X^+ - X^-$, we define $E(X) = E(X^+) - E(X^-)$ provided that at least one of $E(X^+)$ and $E(X^-)$ is finite; if both are infinite then $E(X)$ is undefined. If $E(X)$ is well-defined then

$$E(X) = \int_0^\infty [1 - F(x)]\,dx - \int_{-\infty}^0 F(x)\,dx$$

If X is a continuous random variable and F has pdf f, we have that $dF(x) = f(x)dx$, so the expectation of X is given by

$$E(X) = \int_{-\infty}^\infty x f(x)\,dx.$$

If X is a discrete random variable with pmf $f(x_i) = \Pr(X = x_i), i = 1, 2, ...$, the expectation of X can be expressed as

$$E(X) = \sum_i x_i f(x_i).$$

Since in the discrete case the integral reduces to a summation over the point masses, we will employ the integral representation to denote expectation of a general (continuous or discrete) random variable.

Suppose that X is a random variable with distribution function $F : \mathbb{R} \to [0, 1]$ and inverse (or quantile function) $F^{-1} : [0, 1] \to \mathbb{R}$. Then

$$E(X) = \int_0^1 F^{-1}(t)\,dt$$

if $E(X)$ is well-defined.

If $Y = h(X)$ is a transformation of a random variable X with pdf $f(x)$ then

$$E(Y) = \int_{-\infty}^{\infty} h(x) f(x)\, dx.$$

Example 4.

(1) Suppose $X \sim \text{Bernoulli}(p)$. This is a discrete distribution with $f(0) = 1 - p$ and $f(1) = p$, so $E(X) = 0 \cdot (1-p) + 1 \cdot p = p$. Also, if $Y = X^2$, $E(Y) = 0^2 \cdot (1-p) + 1^2 \cdot p = p$. Observe that in general $E(X^2) \neq E(X)^2$, as it is certainly the case here.

(2) Suppose X follows the Poisson distribution with p.m.f. $f(x) = e^{-\lambda}\lambda^x/x!$, $x = 0, 1, 2, \ldots$.. Using again the Taylor expansion of the exponential given above, it is now difficult to show that

$$E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \lambda.$$

Also, $E(X^2) = \lambda^2 + \lambda$, so that $V(X) = E(X^2) - E(X)^2 = \lambda$, so that λ is both the mean and the variance of the Poisson distribution.

(3) Suppose $X \sim \text{Uniform}[0, 2]$, i.e.,

$$f(x) = \begin{cases} \frac{1}{2}, & \text{for } 0 \leq x \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

We calculate

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^2 \frac{1}{2} x dx = 1.$$

Also, if $Y = X^2$,

$$E(Y) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^2 \frac{1}{2} x^2 dx = \frac{4}{3}.$$

(4) Suppose X follows the standard normal distribution. Then

$$E(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{x^2/2} dx = 0.$$

■

As the integrability condition in the definition of the expectation above indicates, not all random variables have an expectation.

Example 5.

(1) Consider a discrete random variable X with pmf

$$f(x) = \begin{cases} \dfrac{3}{(\pi x)^2}, & \text{for } x = \pm 1, \pm 2, \ldots \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$E(X^+) = \sum_{x=1}^{\infty} x f(x) = \sum_{x=1}^{\infty} \frac{3}{\pi^2 x} = \infty.^3 \tag{5.0}$$

Therefore, $E(X) = \infty - \infty$, and X does not have an expectation.

(2) Let X be a Cauchy random variable. Then

$$E(X) = \int_{-\infty}^{\infty} x \, \frac{1}{\pi(1+x^2)} \, dx = \frac{1}{2\pi} \log(1+x^2) \Big|_{-\infty}^{\infty} = \infty - \infty,$$

and the expectation does not exist. ■

In our examples above we saw that the expectation of the standard normal is 0, but the expectation of the standard Cauchy fails to exist. Why is that?

---

[3]**Theorem.** The series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ diverges for all $p \leq 1$, and converges for all $p > 1$.

Proof: There is a very simple proof of the divergence of the *harmonic series*, that is the series for $p = 1$, that goes as follows. Suppose $\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$ converges to a number S. Then the even numbered terms clearly converge to $\frac{1}{2}S$. But this means that the odd numbered terms must converge to the other half of S, which is impossible because

$$\tfrac{1}{1} > \tfrac{1}{2}, \; \tfrac{1}{3} > \tfrac{1}{4}, \; \tfrac{1}{5} > \tfrac{1}{6}, \ldots .$$

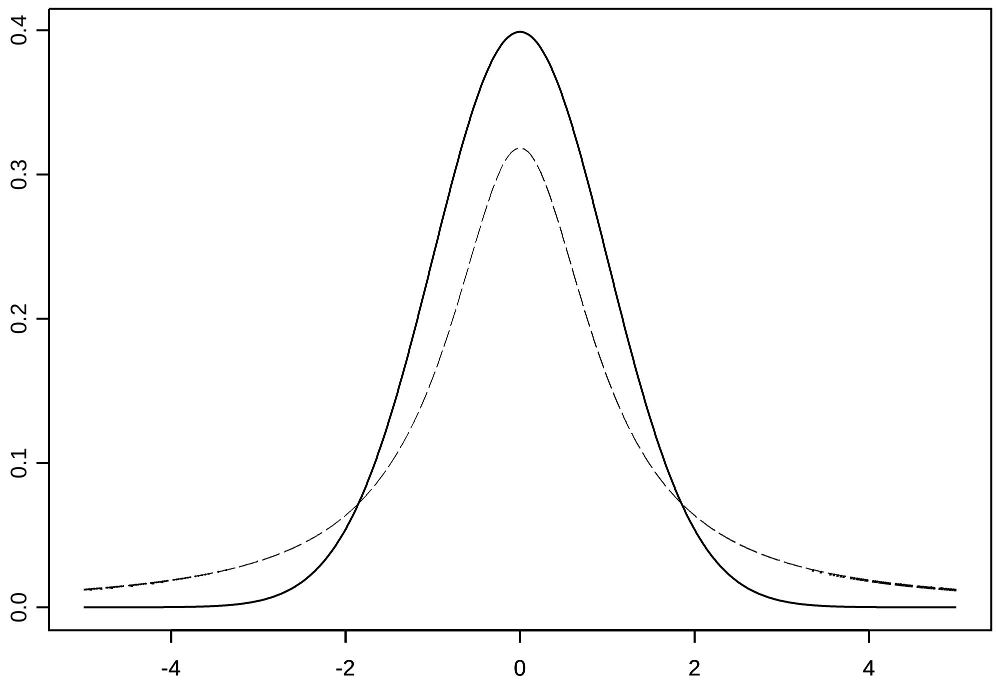
Thus the series must diverge.

Now consider $p < 1$. Since for $p < 1$ the series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ is term by term greater than the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$, it follows immediately that it too must diverge.

Finally consider the more interesting case $p > 1$. Write S_N for the Nth partial sum of the series. Then

$$\begin{aligned} S_{2N+1} &= 1 + \left[\frac{1}{2^p} + \frac{1}{4^p} + \cdots + \frac{1}{(2N)^p}\right] + \left[\frac{1}{3^p} + \frac{1}{5^p} + \cdots + \frac{1}{(2N+1)^p}\right] \\ &< 1 + \left[\frac{1}{2^p} + \frac{1}{4^p} + \cdots + \frac{1}{(2N)^p}\right] + \left[\frac{1}{2^p} + \frac{1}{4^p} + \cdots + \frac{1}{(2N)^p}\right] \\ &= 1 + \frac{1}{2^p} S_N + \frac{1}{2^p} S_N \\ &< 1 + 2^{1-p} S_{2N+1} \end{aligned}$$

because $S_N < S_{2N+1}$. Thus $(1 - 2^{1-p}) S_{2N+1} < 1$. Since $p > 1$ the factor $1 - 2^{1-p}$ is positive, and so we have $S_{2N+1} < (1 - 2^{1-p})^{-1}$ for all N. Since $S_{2N} < S_{2N+1}$ we have that the increasing sequence $\{S_N\}$ is bounded from above by $(1 - 2^{1-p})^{-1}$. Hence it converges. ■



The reason is that the Cauchy density (broken line) has much fatter tails than that of the normal density (solid line), and as it turns out this is exactly the reason for the failure of the Cauchy to have an expectation. The normal is an example of a density with *exponential tails*, while the Cauchy is an example of a density with *algebraic tails*. For scalar random variable X with df F we say F, or X, has an *exponential tail* if

$$\lim_{x\to\infty} \frac{-\log(1-F(x))}{cx^r} = 1, \qquad \text{for some } c>0, r>0,$$

and an *algebraic tail* if

$$\lim_{x\to\infty} \frac{-\log(1-F(x))}{m\log x} = 1, \qquad \text{for some } m>0.$$

Densities with exponential tails die out very fast as we move away from their center, while densities with algebraic tails remain positive even far from their centers. As a result, random samples from densities with algebraic tails tend to have *outliers*, i.e. observations that are far removed from the main body of the data.

Example 6. Use `Mathematica` to compute the limits

$$\lim_{x\to\infty} \frac{-\log(1-\Phi(x))}{x} \quad \text{and} \quad \lim_{x\to\infty} \frac{-\log(1-\Phi(x))}{\log x}$$

to see that the first converges to 1 while the latter diverges to infinity. Thus, the normal has exponential (thin) tails. Also verify that the log-normal has an algebraic right tail. ■

Example 7. The standard logistic cdf

$$F(x) = \frac{e^x}{1+e^x}$$

is frequently described as a distribution with tails somewhat fatter than the normal, but it actually has exponential tails just like the normal. To see this, write

$$\lim_{x\to\infty} \frac{-\log(1-F(x))}{x} = \lim_{x\to\infty} \frac{-\log(1/(1+e^x))}{x} = \lim_{x\to\infty} \frac{\log(1+e^x)}{x},$$

and use L'Hopital's rule to get

$$\lim_{x\to\infty} \frac{\log(1+e^x)}{x} = \lim_{x\to\infty} \frac{d\log(1+e^x)/dx}{dx/dx} = \lim_{x\to\infty} \frac{e^x}{1+e^x} = 1.$$

■

Example 8. Now consider the standard Cauchy cdf. We have

$$F(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi},$$

so

$$\lim_{x\to\infty} \frac{-\log(1-F(x))}{\log x} = \lim_{x\to\infty} \frac{\log(2\pi) - \log(\pi - 2\arctan(x))}{\log x} = 1.$$

so the Cauchy has algebraic (fat) tails. Note that if we divide by x instead of $\log x$ the limit will be zero. The arctan function is *slowly varying at infinity*, i.e., for large x it is strictly increasing but very slowly. The logarithm $\log(x)$ is also slowly varying at infinity so it is exactly what is needed to counter the arctan function and their ratio tends to one. ■

Example 9. The Pareto distribution with pdf

$$f(x) = \begin{cases} 0, & \text{for } x < \beta, \\ \dfrac{\alpha\beta^\alpha}{x^{\alpha+1}}, & \text{for } x \geq \beta \end{cases}$$

and cdf

$$F(x) = \begin{cases} 0, & \text{for } x < \beta, \\ 1 - \left(\dfrac{\beta}{x}\right)^{\alpha}, & \text{for } x \geq \beta \end{cases}$$

is very useful in modeling tail events. Without loss of generality, set $\beta = 1$, in which case

$$\text{logtail}(x) = -\log(1 - F(x)) = \alpha \log x$$

so for $m = \alpha$,

$$\frac{-\log(1 - F(x))}{\alpha \log x} = 1,$$

holds not just in the limit but for *all* x! This means that, at least for large x, *any* distribution with algebraic tails will eventually behave as a Pareto distribution (!), which explains its usefulness in modeling exceedances. For example, if income is log-normally distributed, the upper 10%, say, of incomes (the right tail) will be approximately Pareto distruted. We will return to theses issues at a future lecture when we discuss the theory of extreme events. ■

6. Theorems on Expectations

Theorem 2. *For a random variable X and constants a and b, the random variable $Y = a + bX$ has expectation $E(Y) = a + bE(X)$ and variance $V(Y) = b^2V(X)$.*

Proof: Integrals are linear operators: $E(Y) = \int(a + bx)f(x)dx = a\int f(x)dx + b\int xf(x)dx = a + bE(X)$. ■

Theorem 3. *The variance of a random variable X is equal to the expectation of its square minus the square of its expectation, $V(X) = E(X^2) - E(X)^2$.*

Proof: Let $Y = X - E(X)$ so that $Y^2 = X^2 + E(X)^2 - 2E(X)X$ and apply the previous theorem to get $V(X) = E(Y^2) = E(X^2) + E(X)^2 - 2E(X)^2 = E(X^2) = E(X)^2$. ■

Theorem 4. *Let c be a constant. Then the the mean square error (MSE) of a random variable X about c is $E(X - c)^2 = \sigma^2 + (c - \mu)^2$. When X is an estimator of a population quantity of interest we say that $MSE(X) = VAR(X) + BIAS(X)^2$.*

Proof: Write $(X - c) = (X - \mu) - (c - \mu) = Y - (c - \mu)$. So $(X - c)^2 = Y^2 + (c - \mu)^2 - 2(c - \mu)Y$. Then the first theorem gives $E(X - c)^2 = E(Y) + (c - \mu)^2 - 2(c - \mu)E(Y)$. But $E(Y) = 0$ and $E(Y^2) = \sigma^2$. ■

The following theorem is very important in econometrics. It says that the expected value (mean) of a random variable is the optimal predictor of X under square loss. A *loss function* is any function $L(\cdot)$ satisfying the condition

$$0 < u < v \Rightarrow \begin{cases} 0 &= L(0) \leq L(u) \leq L(v) \\ 0 &= L(0) \leq L(-u) \leq L(-v), \end{cases}$$

where u and v denote values for the *prediction error* $X - c$. The square loss $E(X-c)^2$ and the absolute loss $E|X-c|$ are the two prime examples of loss fuctions.

THEOREM 5. *The value of c that minimizes $E(X-c)^2$ is $c = \mu$, i.e.,*

$$\mu = \underset{c\in\mathbb{R}}{\text{argmin}}\ E(X-c)^2.$$

Proof: From theorem 3, $E(X-c)^2 = \sigma^2 + (c-\mu)^2$, and $(c-\mu)^2 \geq 0$ with equality only when $c = \mu$. ■

7. MOMENTS

The moments of a r.v. X are the expectations of *integer* powers of X, or of $X - \mu$. When $E(X^r)$ exists we call it the r-th *raw moment* of X, or the r-th moment of X around 0, and denote it by m_r, i.e.

$$m_r = E(X^r) = \int x^r dF(x).$$

Clearly then (verify), $m_0 = 1, m_1 = E(X), m_2 = E(X^2)$, and so on.

THEOREM 6. *For integers $r < s$, the existence of m_s implies the existence of m_r.*

Proof: When $|X| \geq 1$, we have $|x|^r \leq |x|^s \Rightarrow \int |x|^r dF(x) \leq \int |x|^s dF(x) < \infty$ and $E(X^r)$ exists. On the other hand, when $0 \leq |X| < 1$, we have $|x|^r \leq |x|^s + 1 \Rightarrow \int |x|^r dF(x) \leq \int 1 + |x|^s dF(x) = 1 + \int |x|^s dF(x) < \infty$, so $E(X^r)$ exists. ■

Next we define the r-th *central moment* of X around the mean $\mu = m_1$ by

$$\mu_r = E(X - m_1)^r = \int (x - m_1)^r dF(x),$$

so that (verify) $\mu_0 = 1, \mu_1 = 0$ and $\mu_2 = m_2 - m_1^2$. The second central moment, μ_2, is called the *variance* of X and is denoted by σ^2.

Theorem 7. .
(a) *If m_r exists, then μ_r also exists and is given by*

$$\mu_r = m_r - \binom{r}{1} m_{r-1} m_1 + \binom{r}{2} m_{r-2} m_1^2 - \cdots + (-1)^{r-1} r m_1^r + (-1)^r m_1^r.$$

(b) *Alternatively, if μ_r exists, then m_r also exists and is given by*

$$m_r = \mu_r + \binom{r}{1} \mu_{r-1} m_1 + \binom{r}{2} \mu_{r-2} m_1^2 + \cdots + r \mu_1 m_1^r + m_1^r.$$

Proof: (Idea) This result follows from the *binomial theorem*: If n a positive integer

$$(a+b)^n = a^n + n a^{n-1} b + \frac{n(n-1)}{2} a^{n-2} b^2 + \cdots + b^n.$$

∎

Apart from the mean $\mu = \mu_1$ and the variance $\sigma^2 = \mu_2$, the *skewness* μ_3/σ^3, and *kurtosis* μ_4/σ^4, are often used measures of *asymmetry* and *tail thickness*, respectively.

The following theorem from Courant (1937)[4], p. 250, says that an improper integral *diverges* if its integrand vanishes at infinity to an order not higher than the first. Since $E(X)$ involves the integrand $xf(x)$, if $f(x)$ vanishes at infinity to an order not higher than the *second* order, $E(X)$ diverges and, *a fortiori*, no higher moments can exist.

Theorem 8. [Courant (1937)]
The improper integral

$$\int_a^\infty f(x)dx, \qquad a > 0, \ f(x) > 0,$$

converges *if the function $f(x)$ vanishes at infinity to a higher order than the first, that is, if there is a number $\nu > 1$ such that for all values of x, no matter how large, the relation*

$$0 < f(x) \leq \frac{M}{x^\nu}$$

is true, where M is a fixed number independent of x. The integral diverges *if the function remains positive and vanishes at infinity to an order not higher than the first, that is, if there there is a fixed number $N > 0$ such that*

$$xf(x) \geq N.$$

[4]Courant, R., *Differential and Integral Calculus.* New York: Interscience Publishers, Inc., Vol. 1, 1937.

In general, if $f(x)$ vanishes at infinity at an order not higher than the $(r+2)$th, the rth and higher moments do not exist. For a random variable to have moments of *all* orders, $f(x)$ must vanish at infinity at an *exponential* rate which is faster than any polynomial. Normal and other well-behaved r.v.'s have this property and posses moments of all orders. It is not difficult to show that for the t_ν r.v., $f(x)$ vanishes at infinity at a rate of $x^{\nu+1}$ so it has only $\nu-1$ finite moments.

8. CONVOLUTIONS

We say that Y is the *convolution* of the independent variables X_1 and X_2, if $Y = X_1 + X_2$. We have

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) = \Pr(X_1 + X_2 \leq y) = \Pr(X_2 \leq y - X_1) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y-x1} f_1(x_1) f_2(x_2) dx_2 dx_1 \\ &= \int_{-\infty}^{\infty} f_1(x_1) \Big[\int_{-\infty}^{y-x1} f_2(x_2) dx_2 \Big] dx_1 \\ &= \int_{-\infty}^{\infty} f_1(x_1) F_2(y - x_1) dx_1. \end{aligned}$$

Hence,

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \int_{-\infty}^{\infty} f_1(x_1) \frac{d}{dy} F_2(y - x_1) dx_1$$

which yields the famous *convolution formula*,

$$f_Y(y) = \int_{-\infty}^{\infty} f_1(x_1) f_2(y - x_1) dx_1.$$

Example 10. (The Tent Distribution). Let X_1 and X_2 be two independent $U[0,1]$ random variables and assume that we wish to find the distribution of $Y = X_1 + X_2$. From it's definition the support of Y is the interval (0,2). Using the convolution formula, we have

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_1(x_1) f_2(y - x_1) dx_1 \\ &= \int_{-\infty}^{\infty} I\{0 \leq x_1 \leq 1\} \cdot I\{0 \leq y - x_1 \leq 1\} dx_1 \\ &= \int_0^1 I\{0 \leq y - x_1 \leq 1\} dx_1. \end{aligned}$$

Consider first $y < 1$. For what values of x_1 would $0 < y - x_1 < 1$? The answer is for $0 < x_1 < y$, so for the $y < 1$ case,

$$f_Y(y) = \int_0^y dx_1 = y, \qquad y < 1.$$

Similarly for $y > 1$, in order to satisfy $0 < y - x_1 < 1$ we need $y - 1 < x_1 < 1$. Thus in this case,

$$f_Y(y) = \int_{y-1}^1 dx_1 = 2 - y, \qquad y > 1.$$

Thus

$$f_Y(y) = \begin{cases} 0 & y \leq 0 \\ y & 0 < y \leq 1 \\ 2 - y & 1 < y \leq 2 \\ 0 & y > 2. \end{cases}$$

For obvious reasons this distribution is called the *tent distribution*. ■

Example 11. Let X_1 and X_2 be two independent $N(0, 1)$ random variables and consider $Y = X_1 + X_2$. We will show that $Y \sim N(0, 2)$. Using the convolution formula,

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_1(x_1) f_2(y - x_1) dx_1 \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\Big\{ -\frac{1}{2}x_1^2 - \frac{1}{2}(y - x_1)^2 \Big\} dx_1. \end{aligned}$$

Completing the square

$$\begin{aligned} x_1^2 + (y - x_1)^2 &= x_1^2 + y^2 - 2x_1 y + x_1^2 \\ &= 2x_1^2 - 2x_1 y + y^2 \\ &= 2\Big(x_1^2 - x_1 y + \frac{1}{4}y^2\Big) + y^2 - \frac{1}{2}y^2 \\ &= 2\Big(x_1 - \frac{y}{2}\Big)^2 + \frac{1}{2}y^2 \end{aligned}$$

we obtain

$$\begin{aligned} f_Y(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\Big\{ -\frac{1}{2}\Big[2\Big(x_1 - \frac{y}{2}\Big)^2 + \frac{1}{2}y^2\Big] \Big\} dx_1 \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\Big\{ -\Big(x_1 - \frac{y}{2}\Big)^2 - \frac{1}{4}y^2 \Big\} dx_1 \\ &= e^{-y^2/4} \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\Big\{ -\Big(x_1 - \frac{y}{2}\Big)^2 \Big\} dx_1 \end{aligned}$$

Multiplying and dividing by $\sqrt{2\pi\frac{1}{2}}$ we obtain

$$f_Y(y) = e^{-y^2/4}\frac{\sqrt{2\pi\frac{1}{2}}}{2\pi}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\frac{1}{2}}}\exp\Big\{-\Big(x_1-\frac{y}{2}\Big)^2\Big\}dx_1.$$

Note that the quantity inside the integral is the density of a $N(\frac{y}{2},\frac{1}{2})$ random variable, so that the integral is equal to 1. Thus

$$f_Y(y) = \frac{1}{2\sqrt{\pi}}e^{-\frac{y^2}{2\cdot 2}} = N(0,2),$$

as promised. ∎

9. Moment Generating Functions

The moment generating function (m.g.f) of a random variable X

$$\psi(t) = E(e^{tX}) = \int e^{tx}dF(x).$$

In other branches of mathematics the m.g.f. is called the *Laplace transform* of a function. Expanding the exponential we obtain,

$$\begin{aligned}\psi(t) &= \int(1+tx+\frac{1}{2}t^2x^2+\cdots)dF(x)\\ &= \sum_{j=0}^{\infty}\int\frac{t^jx^j}{j!}dF(x)\\ &= \sum_{j=0}^{\infty}\frac{t^j}{j!}\int x^jdF(x).\end{aligned}$$

Therefore,

$$\psi(t) = \sum_{j=0}^{\infty}\frac{t^j}{j!}m_j,$$

so that the coefficient of $t^j/j!$ is the j-th raw moment m_j. It follows that one way of finding m_j is by differentiating $\psi(t)$ w.r.t t^j and evaluate at 0, i.e.,

$$m_j = \frac{d^j\psi(t)}{dt^j}\bigg|_{t=0}$$

Example 12.

(1) For the *standard normal* random variable,

$$\begin{aligned}\psi(t) &= E(e^{tX}) \\ &= \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx \\ &= \frac{1}{\sqrt{2\pi}}e^{t^2/2}\int_{-\infty}^{\infty} e^{-(x-t)^2/2}dx.\end{aligned}$$

Let $s = (x - t)$ and change variables to write the integral above as

$$\int_{-\infty}^{\infty} e^{-(x-t)^2/2}dx = \int_{-\infty}^{\infty} e^{-s^2/2}ds = \sqrt{2\pi}.$$

Thus,

$$\psi(t) = e^{t^2/2}.$$

To find the moments of X we expand the mgf to obtain

$$\psi(t) = e^{t^2/2} = \sum_{j=0}^{\infty}\frac{1}{j!}\left(\frac{t^2}{2}\right)^j = \sum_{j=0}^{\infty}\frac{t^{2j}}{(2j)!}\ 1\times 3\times \times 5\times\cdots(2j-1).$$

It follows that all the odd raw moments are zero (by symmetry), while the even raw moments are given by

$$m_{2j} = 1\times 3\times 5\times\cdots\times(2j-1), \quad j = 1, 2, \ldots .$$

(2) For the *Poisson* random variable,

$$\psi(t) = \sum_{x=0}^{\infty} e^{tx}e^{-\lambda}\frac{\lambda^x}{x!} = e^{\lambda(e^t-1)}.$$

We can now compute the moments easily,

$$m_1 = \left.\frac{d\psi(t)}{dt}\right|_{t=0} = \left.\lambda e^t e^{\lambda(e^t-1)}\right|_{t=0} = \lambda$$

and

$$m_2 = \left.\frac{d^2\psi(t)}{dt^2}\right|_{t=0} = \left.(\lambda^2 e^{2t} + \lambda e^t)e^{\lambda(e^t-1)}\right|_{t=0} = \lambda^2 + \lambda,$$

so that,

$$\mu_2 = m_2 - m_1^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Thus, both the mean and the variance of a Poisson random variable is λ.

■

The expansion of the mgf as the weighted sum of an infinity (!) of raw moments given above, rises the question: what if X does not have moments of all orders? As we have seen, some variables don't even have a first moment. Clearly, the m.g.f. would fail to exist if X does not have moments of *all* orders, or to put in another way, if X has an mgf then it also has moments of all orders. It also follows that when the mgf exists, it is unique (given by the above expansion), and completely describes the random variable at hand.

Example 13. Consider a discrete random variable X with pmf

$$f(x) = \frac{6}{(\pi x)^2}, \quad x = 1, 2, 3, \ldots$$

Note that $f(x) = \Pr(X = x) \geq 0$ for all $x = 1, 2, 3, \ldots$ and

$$\sum_{x=1}^{\infty} \frac{6}{(\pi x)^2} = \frac{6}{\pi^2}\Big(\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots\Big) = \frac{6}{\pi^2} \cdot \frac{\pi^2}{6} = 1,$$

so this is a proper pmf. For this random variable,

$$\begin{aligned}
\psi(t) &= E(e^{tX}) \\
&= \frac{6}{\pi^2} \sum_{x=1}^{\infty} e^{tx} \frac{1}{x^2} \\
&= \frac{6}{\pi^2} \sum_{x=1}^{\infty} \Big(1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \cdots\Big) \frac{1}{x^2} \\
&= \frac{6}{\pi^2} \sum_{x=1}^{\infty} \Big(\frac{1}{x^2} + \frac{t}{x} + \frac{t^2}{2!} + \frac{t^3 x}{3!} + \cdots\Big).
\end{aligned}$$

This sum does not converge, and therefore X does not have a moment generating function. Actually, we have already seen that X have no moments at all. ■

It is simple to show that if $\psi_X(t)$ is the mgf of X and $\alpha, \beta \in \mathbb{R}$,

$$\psi_{\alpha+\beta X}(t) = e^{\alpha t} \cdot \psi_X(\beta t).$$

Also if Y is another variable independent of X with mgf $\psi_Y(t)$, the mgf of their sum $X + Y$ is given by

$$\psi_{X+Y}(t) = \psi_X(t) \cdot \psi_Y(t).$$

This last expression for the sum of two random variables is the most useful property of the mgf.

10. Characteristic Functions

The *characteristic function* (cf) of a random variable X is given by,

$$\varphi(t) = E(e^{itX}) = \int e^{itx} dF(x),$$

where $i = \sqrt{-1}$. Unlike the mgf, the cf *always* exists. To see this, recall that $|a + ib| \equiv (a + ib)(a - ib) = a^2 + b^2$, and use the Euler identity $e^{iz} = \cos z + i \sin z$ to write,

$$\begin{aligned} |\varphi(t)| &= \Big| \int e^{itx} dF(x) \Big| \\ &\leq \int |e^{itx}| dF(x) \\ &= \int |\cos tx + i \sin tx| dF(x) \\ &= \int \sqrt{\cos^2 tx + \sin^2 tx}\; dF(x) \\ &= \int 1 dF(x) = 1. \end{aligned}$$

Hence, $|\varphi(t)| \leq 1$, and $\varphi(t)$ always exists. An important property of the cf is that it (essentially) determines its corresponding df through the *inversion formula* for the density

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt,$$

or, more generally,

$$F(x) - F(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-ity}}{it} \varphi(t) dt,$$

if x and y are continuity points of F.

Theorem 9. *If $E|X|^m < \infty$ for an integer $m > 0$, then*

$$\varphi(t) = \sum_{j=0}^{m} \frac{(it)^j}{j!} m_j + o(t^m).$$

Proof: By a Taylor expansion

$$e^{itx} = \sum_{j=0}^{m} \frac{(itx)^j}{j!} + \frac{(it)^m}{(m-1)!} \int_0^1 (e^{itxs} x^m - x^m)(1-s)^{m-1} ds.$$

Now replace x by X and take expectations, evaluating the remainder as,

$$\frac{(it)^m}{(m-1)!} \int_0^1 (\varphi_m(ts) - \varphi_m(0))(1-s)^{m-1} ds.$$

Under the condition that φ_m is uniformly continuous, for small t the integral is $o(1)$ and thus the remainder is $o(t^m)$. ■

When the mgf exists, the cf and the mgf are related by

$$\varphi(t) = \psi(it).$$

More generally, the cf has the same properties as the mgf when the latter exists. The cf of a standard normal random variable is $\varphi(t) = e^{(it)^2/2} = e^{-t^2/2}$, while for a Poisson random variable the cf is $\varphi(t) = e^{\lambda(e^{it}-1)}$. The cf of a Cauchy random variable is given by $e^{-|t|}$ and since this is discontinuous at 0, the Cauchy has no moments.

11. Total Probability and the Generalized Bayes Rule

The *total probability* of an event A whose occurence depends on a r.v. X is given by the formula

$$P(A) = \int_{-\infty}^{\infty} f(x)P(A|x)dx,$$

in which $f(x)$ is the pdf of X, on the values of which depends the probability of occurence of A, and $P(A|x)$ is the probability of A under the assumption that X takes the value x.

The conditional probability density $f(x|A)$ of X under the assumption that A has occured, is determined by the *generalized Bayes formula*:

$$f(x|A) = \frac{f(x)P(A|x)}{\displaystyle\int_{-\infty}^{\infty} f(x)P(A|x)dx}.$$

In this context, $f(x)$ is the *prior* pdf of X (prior to the occurence of A), and $f(x|A)$ is the *posterior* pdf of X.

Example 14. The deviation of the size of an item from the midpoint of the tolerence field of width $2d$ equals the sum of two independent mean zero normal r.v.'s X and Y with variancess σ_x^2 and σ_y^2, respectively. We are interested in determining the conditional probability density of X for the non-defective items.

Let A be the event that an item is non-defective. The conditional probability $P(A|x)$ of getting a non-defective item when X takes the value x is given by

$$P(A|x) = \int_{x-d}^{x+d} \frac{1}{\sigma_y\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) dy = \Phi\left(\frac{x+d}{\sigma_y}\right) - \Phi\left(\frac{x-d}{\sigma_y}\right),$$

where Φ is the standard normal cdf.

Let $f(x|A)$ be the conditional p.d.f of X for non-defective items, so that

$$f(x|A) = \frac{f(x)P(A|x)}{\displaystyle\int_{-\infty}^{\infty} f(x)P(A|x)dx}.$$

Substituting the prior density of X, $f(x) = \phi(x/\sigma_x)$, and $P(A|x)$ from above, we obtain the posterior density of X as

$$\begin{aligned} f(x|A) &= \frac{\dfrac{1}{\sigma_x\sqrt{2\pi}}\exp\left(-\dfrac{x^2}{2\sigma_x^2}\right)\left[\Phi\left(\dfrac{x+d}{\sigma_y}\right) - \Phi\left(\dfrac{x-d}{\sigma_y}\right)\right]}{\displaystyle\int_{-\infty}^{\infty}\dfrac{1}{\sigma_x\sqrt{2\pi}}\exp\left(-\dfrac{x^2}{2\sigma_x^2}\right)\left[\Phi\left(\dfrac{x+d}{\sigma_y}\right) - \Phi\left(\dfrac{x-d}{\sigma_y}\right)\right]dx} \\ &= \frac{\Phi\left(\dfrac{x+d}{\sigma_y}\right) - \Phi\left(\dfrac{x-d}{\sigma_y}\right)}{2\Phi\left(d\Big/\sqrt{\sigma_x^2+\sigma_y^2}\right) - 1} \times \phi(x/\sigma_x), \end{aligned}$$

where ϕ and Φ are the standard normal pdf and cdf, respectively.

Note that as $d/\sigma_y \to 0$, i.e., when only very small deviations relative to the precision of the process (as measured relative to σ_y) are tolerated, the first factor in the last expression tends to $\phi(x/\sigma_y)$ (by the basic definition of the derivative), and

$$f(x|A) \to \phi(x/\sigma_y)\phi(x/\sigma_x) = \phi\left(x/\sqrt{\sigma_x^2\sigma_y^2/(\sigma_x^2+\sigma_y^2)}\right),$$

a rescaled normal pdf. Since when d/σ_y is small the prior and the posterior distributions of X are different, we say that A is *informative* about X .

On the other hand, as d/σ_y becomes large, i.e., as even very large deviations are tolerated, the first factor tends to 1 and

$$f(x|A) \to \phi(x/\sigma_x),$$

the unconditional distribution of X. In this case, we say that the prior is *uninformative*, i.e. knowning that an item is non-defective doesn't tell us much since even items with very large deviations are characterized as "non-defective". ■

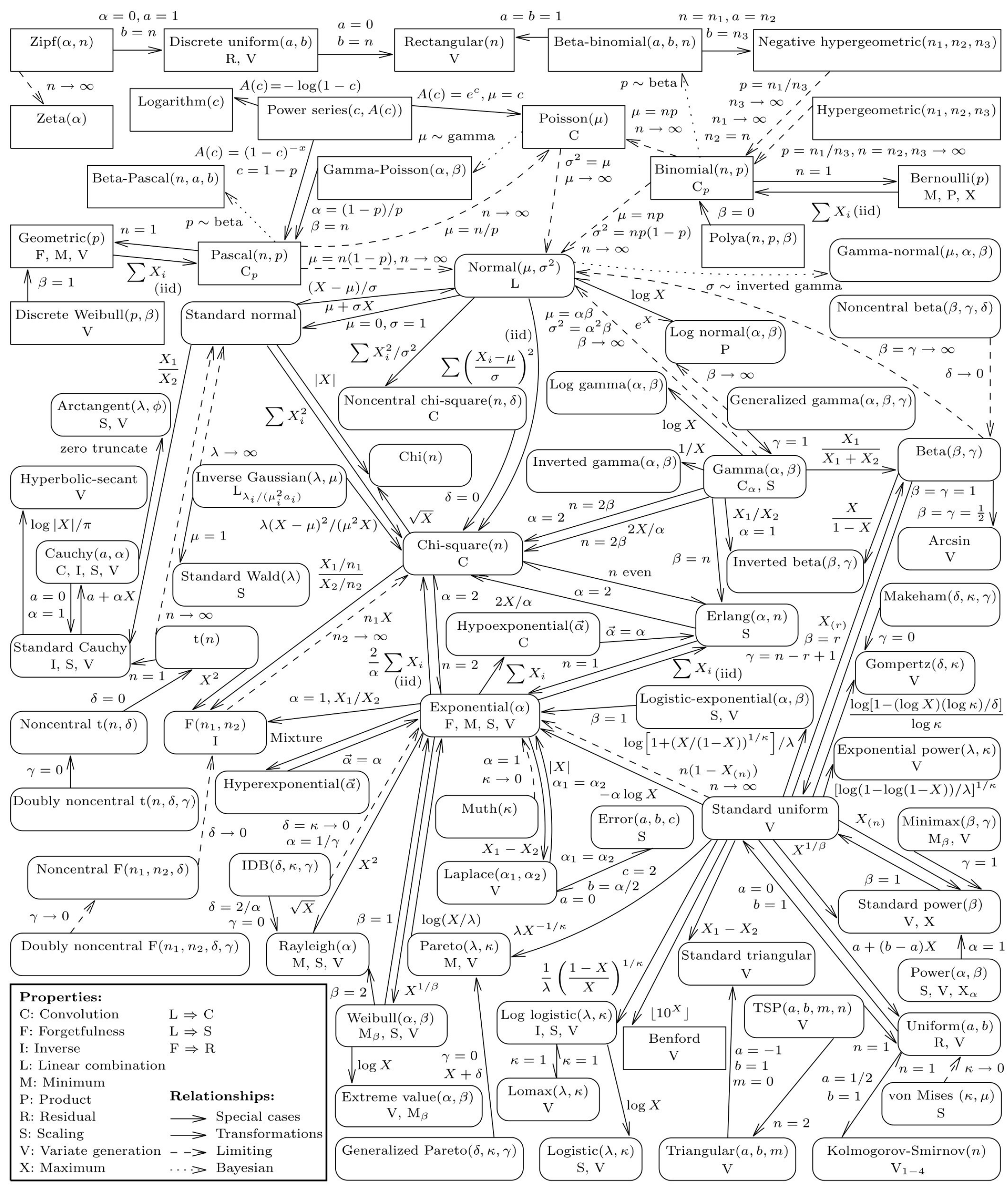


Figure 1. Univariate distribution relationships.

Leemis & McQuestion <http://www.math.wm.edu/~leemis/2008amstat.pdf> *The American Statistician, February 2008, Vol. 62, No. 1* 47

FIGURE 6. Visit http://www.math.wm.edu/ leemis/chart/UDR/UDR.html for an interactive graph.

De Morgan was explaining to an actuary what was the chance that a certain proportion of some group of people would at the end of a given time be alive; and quoted the actuarial formula [the normal density], involving π, which, in answer to a question, he explained stood for the ratio of the circumference of a circle to its diameter. His acquaintance, who had so far listened to the explanation with interest, interrupted him and exclaimed, "My dear friend, that must surely be a delusion, what can a circle have to do with the number of people alive at a given time?"

— *Walter William Rouose Ball* (1850 – 1925)
Mathematical Recreations and Problems,
London: 1896, p.180.