Probability and Statistics Review

Christos Alexopoulos and Dave Goldsman

Georgia Tech

6 Feb. 2020

Outline

Preliminaries

Will assume that you know about sample spaces, events, and the definition of probability.

Definition: If $P(B) > 0$, then $P(A|B) \equiv P(A \cap B)/P(B)$ is the *conditional probability of A given B*.

Example: Toss a fair die. Let $A = \{1, 2, 3\}$ and $B = \{3, 4, 5, 6\}$. Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{4/6} = 1/4.$$

Definition: If $P(A \cap B) = P(A)P(B)$, then A and B are *independent* events.

Theorem: If A and B are independent, then $P(A|B) = P(A)$.

Example: Toss two fair dice. Let $A =$ "Sum is 7" and $B =$ "First die is 4". Then

$$P(A) = 1/6, \quad P(B) = 1/6, \quad \text{and}$$

$$P(A \cap B) = P((4,3)) = 1/36 = P(A)P(B).$$

Thus A and B are independent.

Definition: A *random variable* (RV) X is a function from the sample space Ω to the real line $\mathbb{R}$, i.e., $X : \Omega \to \mathbb{R}$.

Example: Let X be the sum of two dice rolls. Then $X((4,6)) = 10$. In addition,

$$P(X = x) = \begin{cases} 1/36 & \text{if } x = 2 \\ 2/36 & \text{if } x = 3 \\ \vdots & \\ 1/36 & \text{if } x = 12 \\ 0 & \text{otherwise.} \end{cases}$$

Definition: If the number of possible values of a RV X is finite or countably infinite, then X is a *discrete* RV. Its *probability mass function* (pmf) is $f(x) \equiv P(X = x)$. Note that $\sum_x f(x) = 1$.

Example: Flip 2 coins. Let X be the number of heads.

$$f(x) = \begin{cases} 1/4 & \text{if } x = 0 \text{ or } 2 \\ 1/2 & \text{if } x = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Examples: Here are some well-known discrete RVs that you may know: Bernoulli(p), binomial(n, p), geometric(p), negative binomial, Poisson(λ), etc.

Definition: A *continuous* RV is one with probability zero at every individual point. A RV is continuous if there exists a *probability density function* (pdf) $f(x)$ such that $P(X \in A) = \int_A f(x)\,dx$ for every set A. Note that $\int_{\mathbb{R}} f(x)\,dx = 1$.

Example: Pick a random number between 3 and 7. Then

$$f(x) = \begin{cases} 1/4 & \text{if } 3 \le x \le 7 \\ 0 & \text{otherwise.} \end{cases}$$

Examples: Here are some well-known continuous RV's: Uniform(a, b), exponential(λ), Normal(μ, σ^2), etc.

Definition: For any RV X (discrete or continuous), the *cumulative distribution function* (cdf) is defined as

$$F(x) \equiv P(X \le x) = \begin{cases} \sum_{y \le x} f(y) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{x} f(y)\,dy & \text{if } X \text{ is continuous.} \end{cases}$$

Note that $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

Example: Flip two fair coins. Let X be the number of heads.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/4 & \text{if } 0 \le x < 1 \\ 3/4 & \text{if } 1 \le x < 2 \\ 1 & \text{if } x \ge 2. \end{cases}$$

Example: Suppose $X \sim \text{expo}(\lambda)$ (i.e., X has the exponential distribution with parameter $\lambda > 0$). Then $f(x) = \lambda e^{-\lambda x}$, $x \ge 0$, and the cdf is $F(x) = 1 - e^{-\lambda x}$, $x \ge 0$.

Outline

Expected Value

Definition: The *expected value* (or *mean*) of a RV X is

$$\mu \equiv \mathsf{E}[X] \equiv \begin{cases} \sum_x xP(X = x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} xf(x)\,dx & \text{if } X \text{ is continuous.} \end{cases}$$

Example: Suppose that $X \sim \text{Bernoulli}(p)$. Then

$$X = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1 - p \ (= q) \end{cases}$$

and we have $\mathsf{E}[X] = \sum_x xf(x) = p$.

Example: Suppose that $X \sim \mathsf{U}(a, b)$. Then

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

and we have $\mathsf{E}[X] = \int_{-\infty}^{\infty} xf(x)\,dx = (a + b)/2$.

"Law of the Unconscious Statistician": Suppose that $g(X)$ is a proper function of the RV X. Then

$$\mathrm{E}[g(X)] = \begin{cases} \sum_x g(x)f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} g(x)f(x)\,dx & \text{if } X \text{ is continuous.} \end{cases}$$

Example: The discrete RV X has the following pmf:

x	2	3	4
$f(x)$	0.3	0.6	0.1

Then $\mathrm{E}[X^3] = \sum_x x^3 f(x) = 8(0.3) + 27(0.6) + 64(0.1) = 25$.

Example: Suppose $X \sim \mathrm{U}(0,2)$. Then

$$\mathrm{E}[X^n] = \int_{\mathbb{R}} x^n f(x)\,dx = 2^n/(n+1).$$

Definitions: The nth *moment* of X is $\mathsf{E}[X^n]$ and the nth *central moment* of X is $\mathsf{E}[(X - \mathsf{E}[X])^n]$. The *variance* of X is the second central moment:

$$\sigma_X^2 \equiv \mathsf{Var}(X) \equiv \mathsf{E}[(X - \mathsf{E}[X])^2] = \mathsf{E}[X^2] - (\mathsf{E}[X])^2.$$

Example: Suppose $X \sim \text{Bernoulli}(p)$. Recall that $\mathsf{E}[X] = p$. Then

$$\mathsf{E}[X^2] = \sum_x x^2 f(x) = p \quad \text{and}$$

$$\mathsf{Var}(X) = \mathsf{E}[X^2] - (\mathsf{E}[X])^2 = p(1-p) = pq.$$

Example: Suppose $X \sim \mathsf{U}(0, 2)$. By previous examples, $\mathsf{E}[X] = 1$ and $\mathsf{E}[X^2] = 4/3$. So

$$\mathsf{Var}(X) = \mathsf{E}[X^2] - (\mathsf{E}[X])^2 = 1/3.$$

Theorem: $\mathsf{E}[aX + b] = a\mathsf{E}[X] + b$ and $\mathsf{Var}(aX + b) = a^2\mathsf{Var}(X)$.

Definitions: The standard deviation of a RV X is the square root of its variance, that is $\sigma_X \equiv \sqrt{\mathsf{Var}(X)}$.

The coefficient of variation (or relative variation) of X is the ratio $\mathsf{CV}(X) \equiv \sigma_X/\mu$ of the standard deviation to the mean.

It is a *unitless* measure of the relative dispersion of X and is not affected by scaling as $\mathsf{CV}(aX) = \mathsf{CV}(X)$.

Outline

Functions of Random Variables

Problem: Suppose X is a RV with pdf/pmf $f(x)$, and let $Y = h(X)$. Find $g(y)$, the pdf/pmf of Y.

Example: Let X denote the number of H's from two coin tosses. Find the pmf for $Y = X^2 - X$.

x	0	1	2
$f(x)$	1/4	1/2	1/4
$y = x^2 - x$	0	0	2

This implies that $g(0) = P(Y = 0) = P(X = 0 \text{ or } 1) = 3/4$ and $g(2) = P(Y = 2) = 1/4$. In other words,

$$g(y) = \begin{cases} 3/4 & \text{if } y = 0 \\ 1/4 & \text{if } y = 2. \end{cases}$$

Example: Suppose X has pdf $f(x) = |x|$, $-1 \leq x \leq 1$. Find the pdf of $Y = X^2$.

First of all, the cdf of Y is

$$\begin{aligned} G(y) &= P(Y \leq y) \\ &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} |x|\,dx = y, \quad 0 < y < 1. \end{aligned}$$

Thus, the pdf of Y is $g(y) = G'(y) = 1$, $0 < y < 1$, indicating that $Y \sim \mathrm{U}(0, 1)$.

Inverse Transform Theorem: Suppose X is a continuous random variable having cdf $F(x)$. Then, amazingly, $F(X) \sim$ uniform(0,1).

Proof: Assume that $F(x)$ is monotone, and let $Y = F(X)$. Then the cdf of Y is

$$\begin{aligned} P(Y \le y) &= P(F(X) \le y) \\ &= P(X \le F^{-1}(y)) \\ &= F(F^{-1}(y)) = y, \end{aligned}$$

which is the cdf of the U$(0,1)$ distribution.

This result is of fundamental importance when it comes to generating random variates during a simulation.

Example: Suppose $X \sim \text{expo}(\lambda)$, so that its cdf is $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$.

Then the Inverse Transform Theorem implies that

$$F(X) = 1 - e^{-\lambda X} \sim \text{U}(0,1).$$

Now let $U \sim \text{U}(0,1)$ and solve $F(X) = U$ to obtain $X = -\frac{1}{\lambda}\,\ell\text{n}(1-U)$.

After a little algebra, we can also verify that

$$X = -\frac{1}{\lambda}\,\ell\text{n}(U) \sim \text{expo}(\lambda).$$

This is how we can generate realizations from the exponential distribution.

Outline

Joint Distributions

Consider two random variables interacting together, e.g., height and weight.

Definition: The *joint cdf* of X and Y is

$$F(x, y) \equiv P(X \leq x, Y \leq y), \quad \text{for all } x, y.$$

Remark: The *marginal cdf* of X is $F_X(x) = F(x, \infty)$. (We use the X subscript to remind us that it's just the cdf of X all by itself.) Similarly, the *marginal cdf* of Y is $F_Y(y) = F(\infty, y)$.

Definition: If X and Y are discrete, then the *joint pmf* of X and Y is $f(x,y) \equiv P(X=x, Y=y)$. Note that $\sum_x \sum_y f(x,y) = 1$.

Remark: The *marginal pmf* of X is

$$f_X(x) = P(X=x) = \sum_y f(x,y)$$

while the *marginal pmf* of Y is

$$f_Y(y) = P(Y=y) = \sum_x f(x,y).$$

Example: The following table gives the joint pmf $f(x,y)$, along with the respective marginals.

	$X=2$	$X=3$	$X=4$	$f_Y(y)$
$Y=4$	0.3	0.2	0.1	0.6
$Y=6$	0.1	0.2	0.1	0.4
$f_X(x)$	0.4	0.4	0.2	1

Definition: If X and Y are continuous, then the *joint pdf* of X and Y is $f(x,y) \equiv \frac{\partial^2}{\partial x \partial y} F(x,y)$. Note that $\int_{\mathbb{R}} \int_{\mathbb{R}} f(x,y)\,dx\,dy = 1$.

Remark: The *marginal pdf's* of X and Y are

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy \quad \text{and} \quad f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx.$$

Example: Suppose the joint pdf is

$$f(x,y) = \frac{21}{4} x^2 y, \quad x^2 \le y \le 1.$$

Then the marginal pdf's are:

$$f_X(x) = \int_{\mathrm{R}} f(x,y)\,dy = \int_{x^2}^{1} \frac{21}{4} x^2 y\,dy = \frac{21}{8} x^2 (1 - x^4), \quad -1 \le x \le 1$$

and

$$f_Y(y) = \int_{\mathrm{R}} f(x,y)\,dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4} x^2 y\,dx = \frac{7}{2} y^{5/2}, \quad 0 \le y \le 1.$$

Definition: X and Y are *independent* RV's if

$$f(x,y) = f_X(x)f_Y(y) \quad \text{for all } x, y.$$

Theorem: X and Y are independent if we can express their joint pdf/pmf as $f(x,y) = a(x)b(y)$ for some functions $a(x)$ and $b(y)$, and the ranges of x and y where $f(x,y) > 0$ do not depend on each other.

Examples: If $f(x,y) = cxy$ for $0 \le x \le 2$, $0 \le y \le 3$, then X and Y are independent.

If $f(x,y) = \frac{21}{4}x^2y$ for $x^2 \le y \le 1$, then X and Y are *not* independent.

If $f(x,y) = c/(x+y)$ for $1 \le x \le 2$, $1 \le y \le 3$, then X and Y are *not* independent.

Definition: The *conditional pdf* (or *pmf*) of Y given $X = x$ is $f(y|x) \equiv f(x,y)/f_X(x)$.

Example: Suppose $f(x,y) = \frac{21}{4}x^2y$ for $x^2 \le y \le 1$. Then

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{21}{4}x^2y}{\frac{21}{8}x^2(1-x^4)} = \frac{2y}{1-x^4}, \quad x^2 \le y \le 1.$$

Theorem: If X and Y are independent, then $f(y|x) = f_Y(y)$ for all x, y.

"(Multivariate) Law of the Unconscious Statistician": Suppose that $h(X, Y)$ is a function of the RVs X and Y. Then

$$\mathsf{E}[h(X,Y)] = \begin{cases} \sum_x \sum_y h(x,y) f(x,y) & \text{if } (X,Y) \text{ is discrete} \\ \int_{\mathbb{R}} \int_{\mathbb{R}} h(x,y) f(x,y)\, dx\, dy & \text{if } (X,Y) \text{ is continuous} \end{cases}$$

Theorem: Whether or not X and Y are independent, we have $\mathsf{E}[X+Y] = \mathsf{E}[X] + \mathsf{E}[Y]$.

Theorem: If X and Y are *independent*, then $\mathsf{E}[XY] = \mathsf{E}[X]\mathsf{E}[Y]$ and $\mathsf{Var}(X+Y) = \mathsf{Var}(X) + \mathsf{Var}(Y)$.

(Stay tuned for dependent RVs . . . .)

Definition: $X_1, \ldots, X_n$ form a *random sample* from $f(x)$ if (i) $X_1, \ldots, X_n$ are independent, and (ii) each X_i has the same pdf/pmf $f(x)$.

Notation: $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} f(x)$. (The term "iid" reads *independent and identically distributed.*)

Example: If $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} f(x)$ and the *sample mean* $\bar{X}_n \equiv \sum_{i=1}^n X_i/n$, then $\mathsf{E}[\bar{X}_n] = \mathsf{E}[X_i]$ and $\mathsf{Var}(\bar{X}_n) = \mathsf{Var}(X_i)/n$. Thus, the variance *decreases* at rate $1/n$ as n increases.

Covariance and Correlation

Definition: The *covariance* between X and Y is

$$\mathsf{Cov}(X,Y) \equiv \mathsf{E}[(X-\mathsf{E}[X])(Y-\mathsf{E}[Y])] = \mathsf{E}[XY]-\mathsf{E}[X]\mathsf{E}[Y].$$

Note that $\mathsf{Var}(X) = \mathsf{Cov}(X,X)$.

Theorem: If X and Y are independent RVs, then $\mathsf{Cov}(X,Y) = 0$.

Remark: $\mathsf{Cov}(X,Y) = 0$ doesn't mean X and Y are independent!

Example: Suppose $X \sim \text{uniform}(-1,1)$ and $Y = X^2$. Then X and Y are clearly dependent. However,

$$\mathsf{Cov}(X,Y) = \mathsf{E}[X^3]-\mathsf{E}[X]\mathsf{E}[X^2] = \mathsf{E}[X^3] = \int_{-1}^{1} \frac{x^3}{2}\,dx = 0.$$

Theorem: $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$.

Theorem: We have

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

and

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y).$$

Definition: The *correlation* between X and Y is

$$\rho \equiv \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Theorem: $-1 \le \rho \le 1$.

Example: Consider the following joint pmf.

$f(x,y)$	$X=2$	$X=3$	$X=4$	$f_Y(y)$
$Y=40$	0.00	0.20	0.10	0.3
$Y=50$	0.15	0.10	0.05	0.3
$Y=60$	0.30	0.00	0.10	0.4
$f_X(x)$	0.45	0.30	0.25	1

We have $\mathsf{E}[X] = 2.8$, $\mathsf{Var}(X) = 0.66$, $\mathsf{E}[Y] = 51$, $\mathsf{Var}(Y) = 69$,

$$\mathsf{E}[XY] = \sum_x \sum_y xyf(x,y) = 140,$$

and

$$\rho = \frac{\mathsf{E}[XY] - \mathsf{E}[X]\mathsf{E}[Y]}{\sqrt{\mathsf{Var}(X)\mathsf{Var}(Y)}} = -0.415.$$

Outline

Bernoulli and Binomial Distributions

A Bernoulli(p) RV X has pmf

$$f(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p \ (= q) & \text{if } x = 0. \end{cases}$$

Then $\mathsf{E}[X] = p$ and $\mathsf{Var}(X) = pq$.

$Y \sim$ binomial(n, p) if it counts is the number of successes in n independent Bernoulli(p) trials. Its pmf is

$$f(y) = \binom{n}{y} p^y q^{n-y}, \quad y = 0, 1, \ldots, n.$$

Further, $\mathsf{E}[Y] = np$ and $\mathsf{Var}(Y) = npq$.

Alternatively, $Y = \sum_{i=1}^{n} X_i$, where $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim}$ Bernoulli(p).

Geometric Distribution

$X \sim$ Geometric(p) if it counts independent Bernoulli(p) trials until the first success occurs. For example, the outcome "FFFS" implies that $X = 4$. Clearly,

$$f(x) = q^{x-1} p, \quad x = 1, 2, \ldots.$$

It turns out that $\mathsf{E}[X] = 1/p$ and $\mathsf{Var}(X) = q/p^2$.

Remark: The random variable $Y = X - 1$ also has the geometric distribution with pmf

$$f(y) = q^y p, \quad y = 0, 1, \ldots,$$

$\mathsf{E}[Y] = q/p$ and $\mathsf{Var}(Y) = q/p^2$.

Fact: The geometric is the only discrete distribution with the "memoryless" property.

Negative Binomial Distribution

A negative binomial random variable Y is the sum of r iid geometric(p) RVs, i.e., the number of Bernoulli trials until the rth success occurs. For example, the outcome "FFFSSFS" implies that $Y = 7$. Since there are $r-1$ successes in the first $y-1$ trials, the pmf of Y is

$$f(y) = \binom{y-1}{r-1} q^{y-r} p^r, \quad y = r, r+1, \ldots.$$

Further, $\mathsf{E}[Y] = r/p$ and $\mathsf{Var}(Y) = rq/p^2$. Notice that the number of trials between consecutive successes is geometric(p).

Remark: The random variable $Z = Y - r$ that counts the number of failures until the rth success also has the negative binomial distribution with pmf

$$f(z) = \binom{r+z-1}{r-1} q^z p^r, \quad z = 0, 1, \ldots,$$

$\mathsf{E}[Z] = rq/p$,and $\mathsf{Var}(Z) = rq/p^2$.

A More Flexible Negative Binomial Distribution

A more flexible negative binomial distribution, frequently used to model demand sizes, allows the parameter $r > 0$ to be noninteger. The pmf of Z is given by

$$f(z) = \frac{\Gamma(r+z)}{\Gamma(r)z!} q^z p^r, \quad z = 0, 1, \ldots,$$

where $\Gamma(\alpha) \equiv \int_0^\infty e^{-t} t^{\alpha-1}\, dt$, $\alpha > 0$ is the gamma function.

It turns out that $\mathsf{E}[Z] = rq/p$ and $\mathsf{Var}(Z) = rq/p^2$. When p is small, the coefficient of variation of Z

$$\mathsf{CV}(Z) = \frac{\sqrt{\mathsf{Var}(Z)}}{\mathsf{E}[Z]} = \frac{1}{\sqrt{rq}}$$

can be larger than 1 when $r < 1/q$.

Fact: The gamma function satisfies the recursion $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$, for $\alpha > 1$. If $\alpha =$ integer, then $\Gamma(\alpha) = (\alpha - 1)!$.

The Poisson Distribution

The Poisson(λ) distribution ($\lambda > 0$) models counts events in a time interval (or space) when the occurrence of an event is independent of the time elapsed since the previous event.

Examples are the number of arrivals in a given time interval, the number of accidents in a power plant during a certain time window, etc.

The pmf of $X \sim$ Poisson(λ) is

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots.$$

After some algebra we can show that $\mathrm{E}[X] = \lambda = \mathrm{Var}(X)$.

Uniform and Triangular Distributions

We proceed with some continuous distributions...

The $\mathsf{U}(a,b)$ distribution has density

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise,} \end{cases}$$

$\mathsf{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx = (a+b)/2$ and $\mathsf{Var}(X) = (b-a)^2/12$.

Further,

$$X \sim \mathsf{U}(a,b) \iff (X-a)/(b-a) \sim \mathsf{U}(0,1).$$

The triangular(a, b, c) distribution is a reasonable model in the presence of limited data. a is the smallest possible value, b is the "most likely" value (mode), and c is the largest possible value. The density function is

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & \text{if } a < x \le b \\ \frac{2(c-x)}{(c-b)(c-a)} & \text{if } b < x \le c \\ 0 & \text{otherwise.} \end{cases}$$

After some algebra, we can show that $\mathsf{E}[X] = (a + b + c)/3$ and $\mathsf{Var}(X) = (a^2 + b^2 + c^2 - ab - bc - ac)/18$.

Fact: Did you know that the sum of two iid $\mathsf{U}(0, 1)$ RVs has the triangular$(0, 1, 2)$ distribution?

Beta Distribution

A more flexible model than the triangular distribution is the beta(α, β) model with density

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1,$$

where $\alpha > 0$ and $\beta > 0$ are "shape" parameters.

It turns out that

$$E[X] = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad \mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Mirror image property: $X \sim \text{beta}(\alpha, \beta) \Leftrightarrow 1 - X \sim \text{beta}(\beta, \alpha)$.

The Pert(a, b, c) distribution used in Simio (also denoted as beta-Pert) with minimum value $a > 0$, most likely value b, and maximum value c is obtained from the following transformation (the value in <> is omitted):

$$X \sim \text{beta}\left(1+4\,\frac{b-a}{c-a}, 1+4\,\frac{c-b}{c-a}\right) \Longleftrightarrow a+(c-a)X \sim \text{Pert}(a, b, c, <4>).$$

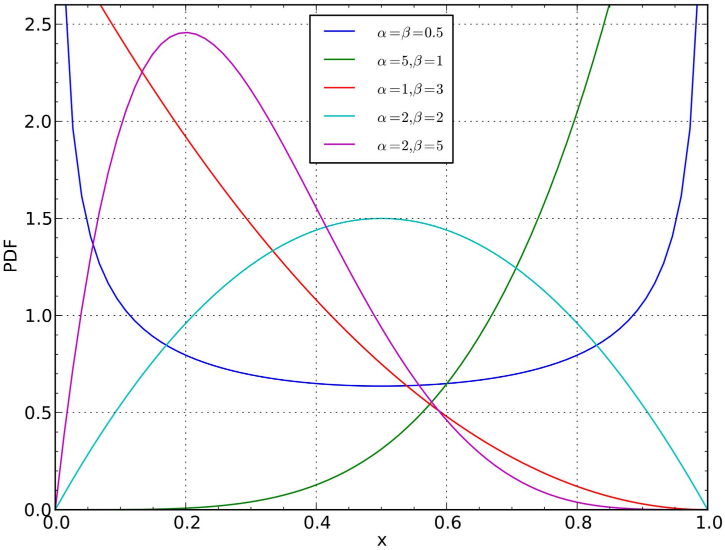

Figure: Plots of beta densities

Exponential Distribution

The expo(λ) distribution has density

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0,$$

cdf

$$F(x) = 1 - e^{-\lambda x}, \quad x > 0,$$

$\mathsf{E}[X] = 1/\lambda$, and $\mathsf{Var}(X) = 1/\lambda^2$.

Theorem: The exponential distribution is the only continuous distribution with the memoryless property:

$$P(X > s + t | X > s) = P(X > t) \quad \text{for } s, t > 0.$$

Example: Suppose $X \sim \text{expo}(\lambda = 1/100)$. Then

$$P(X > 200 | X > 50) = P(X > 150) = e^{-\lambda t} = e^{-150/100} = e^{-1.5} = 0.223.$$

Gamma Distribution

The gamma(α, λ) distribution with shape parameter $\alpha > 0$ and scale parameter $\lambda > 0$ has density function

$$f(x) = \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad x \geq 0,$$

where the gamma function was defined earlier.

It turns out that $\mathsf{E}[X] = \alpha/\lambda$ and $\mathsf{Var}(X) = \alpha/\lambda^2$; hence $\mathsf{CV}(X) = 1/\sqrt{\alpha}$.

Gamma Distribution (cont'd)

Facts:

- When $\alpha = 1$, the gamma(α, λ) distribution reduces to expo(λ).
- $X \sim \text{gamma}(\alpha, \lambda) \iff \lambda X \sim \text{gamma}(\alpha, 1)$.
- Excel and Simio use the notation gamma(α, β) with $\beta = 1/\lambda$.
- If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{expo}(\lambda)$, then $X \equiv \sum_{i=1}^{n} X_i \sim \text{gamma}(n, \lambda)$. The gamma$(n, \lambda)$ distribution is also denoted as Erlang(n, λ) and has cdf

$$F(x) = 1 - e^{-\lambda x} \sum_{j=0}^{n-1} \frac{(\lambda x)^j}{j!}, \quad x \geq 0.$$

(Stay tuned for the proof...)

Gamma Distribution (cont'd)

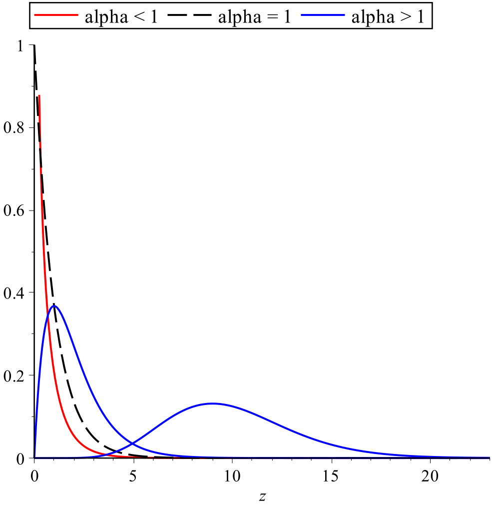

Figure: Plots of gamma densities with $\lambda = 1$

Poisson Process (PP)

The (stationary) Poisson process counts events (e.g., entity arrivals) in time intervals.

Let $N(t)$ tally the number of events observed in $[0, t]$.

Definition: We say that $\{N(t) : t \geq 0\}$ is a Poisson process with rate λ if the times between successive events are iid expo(λ). Equivalently,

(a) events occur one-at-a-time at rate λ;

(b) the increments $N(s+t) - N(s)$ for $s, t \geq 0$ are independent, i.e., the event counts in disjoint time intervals are independent;

(c) the increments are stationary, i.e., the distribution of the number of events in $[s, s+t]$ only depends on t.

Under these assumptions we can show that

$$N(s+t) - N(s) \sim \text{Poisson}(\lambda t);$$

hence the expected rate of events is $\mathsf{E}[N(s+t) - N(s)]/t = \lambda$ (constant).

Poisson Process (cont'd)

Let X_i between events i and $i+1$, and let $S_n = \sum_{i=1}^{n} X_i$.

Facts:

- $X_i \overset{\text{iid}}{\sim} \text{expo}(\lambda)$.
- $S_n \sim \text{gamma}(n, \lambda)$ and

$$
\begin{aligned}
P(S_n \le t) &= P\{\text{at least } n \text{ events in } [0,t]\} \\
&= 1 - P\{\text{at most } n-1 \text{ events in } [0,t]\} \\
&= 1 - P\{N(t) \le n-1\} \\
&= 1 - \sum_{j=0}^{n-1} e^{-\lambda t} \frac{(\lambda t)^j}{j!} \\
&= 1 - e^{-\lambda t} \sum_{j=0}^{n-1} \frac{(\lambda t)^j}{j!}.
\end{aligned}
$$

Example

At a nuclear plant, (minor) accidents occur according to a Poisson process with a rate of one every two years.

- What is the probability that the time between successive accidents is greater than two years?
  Answer: The rate of accidents is $\lambda = 1/2$ per year. Hence the requested probability is $e^{-(1/2)2} = e^{-1} = 0.368$.
- What is the probability that 4 or more accidents will occur in a two-year interval?
  Answer: $N(2) \sim \text{Poisson}(1)$ and
  $P\{(N(2) \geq 4\} = 1 - e^{-1}(1 + 1 + 1/2 + 1/6) = 0.019$.
- What is the mean number of accidents in ten years?
  Answer: $(1/2)10 = 5$.
- Suppose that a year has passed since the last accident. What is the probability that the next accident will occur at least three years from now?
  Answer: By the memoryless property, the answer is $e^{-(1/2)3} = 0.223$.

Nonstationary Poisson Process (NPP)

Such processes have independent but nonstationary increments. There is a rate function $\lambda(t) \geq 0$ with a cumulative function $\Lambda(t) = \int_0^t \lambda(s)\,ds$ such that the increments have the Poisson distribution with a mean that depends on the location and length of the respective time interval:

$$N(s+t) - N(s) \sim \text{Poisson}\left(\int_s^{s+t} \lambda(u)\,du\right) = \text{Poisson}[\Lambda(s+t) - \Lambda(s)].$$

Connection Between PPs and NPPs:

- Let $T_1 < T_2 < \cdots$ be the event times in an NPP with rate function $\lambda(t)$. Then the $\tau_i \equiv \Lambda(T_i)$ are event times in a PP with rate 1.
- Conversely, let $\tau_1 < \tau_2 < \cdots$ be the event times in a PP with rate 1. Then $\Lambda^{-1}(\tau_i)$ are event times in a NPP with rate function $\lambda(t)$.

Example

Customers arrive at a Post Office as an NPP with rates of 2 per minute between 8 a.m. and 12 p.m., and then 0.5 per minute until 4 p.m. Let $t = 0$ correspond to 8 a.m.. The NPP $\{N(t)\}$ has rate function

$$\lambda(t) = \begin{cases} 2 & \text{for } 0 \le t < 4 \\ 0.5 & \text{for } 4 \le t \le 8. \end{cases}$$

The expected number of arrivals by time t is given by the cumulative rate function

$$\Lambda(t) = \begin{cases} 2t, & \text{for } 0 \le t < 4 \\ \int_0^4 2\,du + \int_4^t 0.5\,du = \frac{t}{2} + 6 & \text{for } 4 \le t \le 8. \end{cases}$$

The distribution of the number of arrivals between 11 a.m. and 2 p.m. is Poisson with mean $\Lambda(6) - \Lambda(3) = 3$.

Weibull Distribution

The Weibull(α, λ) distribution with shape parameter $\alpha > 0$ and scale parameter $\lambda > 0$ has density function

$$f(x) = \alpha\lambda(\lambda x)^{\alpha-1}e^{-(\lambda x)^{\alpha}}, \quad x > 0$$

and cdf

$$F(x) = 1 - e^{-(\lambda x)^{\alpha}}, \quad x > 0.$$

It turns out that $\mathsf{E}[X] = \frac{1}{\lambda}\Gamma(1 + \frac{1}{\alpha})$.

Facts:

- When $\alpha = 1$, the Weibull(α, λ) distribution reduces to expo(λ).
- $X \sim$ Weibull(α, λ) $\Leftrightarrow \lambda X \sim$ Weibull($\alpha, 1$).
- Excel and Simio use the notation Weibull(α, β) with $\beta = 1/\lambda$.

Weibull Distribution (cont'd)

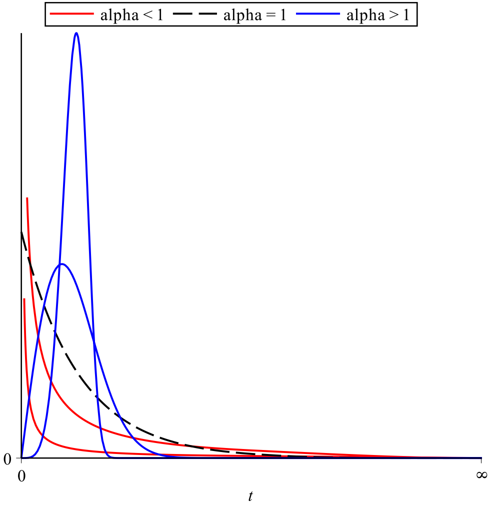

Figure: Plots of Weibull densities with $\lambda = 1$

Normal Distribution

The $\mathrm{N}(\mu, \sigma^2)$ is the most important distribution in probability and statistics. It has density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \quad x \in \mathbb{R},$$

$\mathrm{E}[X] = \mu$, and $\mathrm{Var}(X) = \sigma^2$.

Theorem: If $X \sim \mathrm{N}(\mu, \sigma^2)$, then $aX + b \sim \mathrm{N}(a\mu + b, a^2\sigma^2)$.
It follows that if $X \sim \mathrm{N}(\mu, \sigma^2)$, then $Z \equiv \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$, the *standard normal distribution*, with cdf $\Phi(z)$, which is tabulated. E.g., $\Phi(1.96) \doteq 0.975$.

Theorem: If X_1 and X_2 are *independent* with $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, $i = 1, 2$, then $X_1 + X_2 \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Example: Suppose $X \sim \mathrm{N}(3,4)$, $Y \sim \mathrm{N}(4,6)$, and X and Y are independent. Then $2X - 3Y + 1 \sim \mathrm{N}(-5, 70)$.

Sampling Distributions

There are a number of distributions (including the normal) that come up in statistical sampling problems. Here are a few:

Definitions: If $Z_1, Z_2, \ldots, Z_k$ are iid $\mathrm{N}(0,1)$, then $Y = \sum_{i=1}^{k} Z_i^2$ has the chi square distribution with k degrees of freedom (df). We write $Y \sim \chi^2_k$. Note that $\mathrm{E}[Y] = k$ and $\mathrm{Var}(Y) = 2k$.

If $Z \sim \mathrm{N}(0,1)$ and $Y \sim \chi^2_k$ are independent, then $T = Z/\sqrt{Y/k}$ has Student's t distribution with k df. We write $T \sim t_k$.

If $Y_1 \sim \chi^2_m$ and $Y_2 \sim \chi^2_n$ are independent, then $F = (Y_1/m)/(Y_2/n)$ has the F distribution with m and n df. We write $F \sim F_{m,n}$.

Outline

Limit Theorems

Corollary (of theorem from previous section): If $X_1, \ldots, X_n$ are iid $\mathsf{N}(\mu, \sigma^2)$, then the sample mean $\bar{X}_n \sim \mathsf{N}(\mu, \sigma^2/n)$.

This is a special case of the *Law of Large Numbers*, which says that $\bar{X}_n$ converges to μ in probability as $n \to \infty$.

Definition: A sequence of RVs $\{X_1, X_2, \ldots\}$ with respective cdf's $F_{X_1}(x), F_{X_2}(x), \ldots$ *converges in distribution* to the RV X having cdf $F_X(x)$ if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for all x where the limiting cdf $F_X(x)$ is continuous. We write $X_n \xrightarrow{d} X$.

Idea: If $X_n \xrightarrow{d} X$ and n is large, then we ought to be able to approximate the distribution of X_n by the limiting distribution of X.

Central Limit Theorem: If $X_1, X_2, \ldots, X_n$ are iid with mean μ and variance $\sigma^2 < \infty$, then

$$Z_n \equiv \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathrm{N}(0,1).$$

Thus, the cdf of Z_n approaches $\Phi(z)$ as n increases. The CLT usually works well if the pdf/pmf is fairly symmetric and $n \geq 20$.

Example: Suppose $X_1, X_2, \ldots, X_{100} \overset{\text{iid}}{\sim} \text{expo}(1)$ (so $\mu = \sigma^2 = 1$).

$$\begin{aligned} P\left(90 \leq \sum_{i=1}^{100} X_i \leq 110\right) &= P\left(\frac{90-100}{\sqrt{100}} \leq Z_{100} \leq \frac{110-100}{\sqrt{100}}\right) \\ &\approx P(-1 \leq \mathrm{N}(0,1) \leq 1) \\ &= 2\Phi(1) - 1 = 0.683. \end{aligned}$$

Outline

Statistics Tidbits

For now, suppose that $X_1, X_2, \ldots, X_n$ are iid from some distribution with finite mean μ and finite variance σ^2.

In this case, we have already seen that $\mathsf{E}[\bar{X}_n] = \mu$, i.e., $\bar{X}_n$ is *unbiased* for μ.

Definition: The *sample variance* is $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$.

Theorem: If $X_1, X_2, \ldots, X_n$ are iid with variance σ^2, then $\mathsf{E}[S_n^2] = \sigma^2$, i.e., S_n^2 is unbiased for σ^2.

In particular, when the X_i are $\mathsf{N}(\mu, \sigma^2)$, then $\bar{X}_n \sim \mathsf{N}(\mu, \sigma^2/n)$, $S_n^2 \sim \frac{\sigma^2 \chi^2_{n-1}}{n-1}$, and $\bar{X}_n$ and S_n^2 are independent.

These facts can be used to construct *confidence intervals* (CIs) for μ and σ^2 under a variety of assumptions.

A $100(1-\alpha)\%$ two-sided CI for an unknown parameter θ is a random interval $[L, U]$ such that $P(L \le \theta \le U) = 1 - \alpha$.

Here are some results, all of which assume that the X_i are iid normal...

Example: If σ^2 is *known*, then a $100(1-\alpha)\%$ CI for μ is

$$\bar{X}_n - z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}} \;\le\; \mu \;\le\; \bar{X}_n + z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}},$$

where $z_{1-\gamma}$ is the $1-\gamma$ quantile of the standard normal distribution, i.e., $z_{1-\gamma} \equiv \Phi^{-1}(1-\gamma)$.

Example: If σ^2 is *unknown*, then a $100(1-\alpha)\%$ CI for μ is

$$\bar{X}_n - t_{1-\alpha/2,n-1}\sqrt{\frac{S_n^2}{n}} \;\le\; \mu \;\le\; \bar{X}_n + t_{1-\alpha/2,n-1}\sqrt{\frac{S_n^2}{n}},$$

where $t_{\nu,1-\gamma}$ is the $1-\gamma$ quantile of the t_ν distribution.

Meaning of this CI: If we repeat this sampling experiment many times, each time with n data points, the fraction of the CIs that contain the true mean μ will be close to $1-\alpha$.

Example: A $100(1-\alpha)\%$ CI for σ^2 is

$$\frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha/2}} \;\le\; \sigma^2 \;\le\; \frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha/2}},$$

where $\chi^2_{\nu,1-\gamma}$ is the $1-\gamma$ quantile of the χ^2_ν distribution.