

Ma 3/103 Introduction to Probability and Statistics KC Border Winter 2020

Supplement 2: Review Your Distributions

Relevant textbook passages:

Pitman [10]: pages 476–487.

Larsen–Marx [9]: Chapter 4

A terrific reference for the familiar distributions is the compact monograph by Forbes, Evans, Hastings, and Peacock [4]. There are also the classics by Johnson and Kotz [6, 7, 8]. In addition, Wikipedia generally has a helpful page devoted to each family of distributions.

S2.1 Bernoulli

The Bernoulli distribution is a discrete distribution that generalizes coin tossing.

A Bernoulli(p) random variable X takes on two values: 1 ("success"), with probability p, and 0 ("failure"), with probability 1 - p. The probability mass function is

$$p(X = x) = \begin{cases} p & x = 1, \\ 1 - p & x = 0. \end{cases}$$

The Bernoulli(p) has mean

$$E X = \sum_{x=0,1} xp(X=x) = 0(1-p) + 1p = p.$$

Moreover

$$E X^2 = \sum_{x=0,1} x^2 p(X=x) = 0^2 (1-p) + 1^2 p = p,$$

so the variance is

$$Var X = E(X^2) - (E X)^2 = p - p^2 = p(1 - p).$$

Note that the variance is maximized for

$$p = 1/2.$$

Also note that every moment is the same:

$$E X^{\alpha} = 0^{\alpha} (1-p) + 1^{\alpha} p = p.$$





S2.2 Rademacher

The Rademacher(p) distribution is a recoding of the Bernoulli distribution, 1 still indicates success, but failure is coded as -1.

If Y is a Bernoulli(p) random variable, then X = 2Y - 1 is a Rademacher(p) random variable.

The probability mass function is

$$P(X = x) = \begin{cases} p & x = 1, \\ 1 - p & x = -1. \end{cases}$$

A Rademacher(p) random variable X has mean

$$E X = \sum_{x=0,1} xp(X=x) = -1(1-p) + 1p = 2p - 1.$$

Moreover $X^2 = 1$ so

$$\boldsymbol{E}(X^2) = 1$$

so the variance is

$$Var X = E(X^2) - (EX)^2 = 1 - (2p - 1)^2 = 4p(1 - p).$$

A sequence of successive sums of independent Rademacher(1/2) random variables is called a **random walk**.

That is, if X_i are independent and identically distributed Rademacher(1/2) random variables, the sequence S_1, S_2, \ldots is a random walk, where

$$S_n = X_1 + \dots + X_n.$$

Since expectation is a linear operator,

$$\boldsymbol{E} S_n = 0$$
 for all n ,

and since the variance of a sum of independent random variables is the sum of the variances,

$$Var S_n = n$$
 for all n .

S2.3 Binomial

The Binomial(n, p) is the distribution of the number X of successes in n independent Bernoulli(p) trials.

It is the sum of n independent Bernoulli(p) random variables.

The probability mass function is

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Since expectation is a linear operator, the expectation of a Binomial is the sum of the expectations of the underlying Bernoullis, so if X has a Binomial(n, p) distribution,

$$\boldsymbol{E} X = np,$$

and since the variance of the sum of independent random variables is the sum of the variances,

$$Var X = np(1-p).$$





S2.4 The Multinomial Distribution

The multinomial distribution generalizes the binomial distribution to independent random experiments with more than two outcomes.

Marx [9]: Section 10.2, pp. 494–499 Pitman [10]: p. 155

Larsen-

A multinomial random vector is an m-vector X of counts of outcomes in a sequence of n independent repetitions of a random experiment with m distinct outcomes.

If the experiment has m possible outcomes, and the i^{th} outcome has probability p_i , then the Multinomial(n, p) probability mass function is given by

$$P(\mathbf{X} = (k_1, \dots, k_m)) = \frac{n!}{k_1! \cdot k_2! \cdots k_m!} p_1^{k_1} \cdot p_2^{k_2} \cdots p_m^{k_m}.$$

where $k_1 + \cdots + k_m = n$.

If we count outcome k as a "success," then it is obvious that each X_k is simply a Binomial (n, p_k) random variable. But the components are not independent, since they sum to n.

S2.5 The Negative Binomial Distribution

Replicate a Bernoulli(p) experiment independently until the r^{th} success occurs ($r \ge 1$). Let X be the number of the trial on which the r^{th} success occurs. Then the random variable X is said to have a **Negative Binomial**(r, p) **Distribution**.

Warning: There is another definition of the Negative Binomial Distribution, namely the distribution of the number of failures before the r^{th} success. (This is the definition employed by Jacod and Protter [5, p. 31], R, and MATHEMATICA.)

The relationship between the two is quite simple. If X is a Negative Binomial(r, p) random variable in the Pitman–Larsen–Marx sense, and F is negative binomial in the R–MATHEMATICA sense, then

$$F = X - r.$$

A simple way to tell which definition your software is using is to ask it the probability that the equals zero. For PLM, it is always 0, but for RM, it is p^r .

What is the probability that the r^{th} success occurs on trial t, for $t \ge r$? For this to happen, there must be t - r failures and r - 1 successes in the first t - 1 trials, with a success on trial t. By independence, this happens with the binomial probability for r - 1 successes on t - 1 trials times the probability p of success on trial t:

$$P(X = t) = {\binom{t-1}{r-1}} p^r (1-p)^{t-r} \qquad (t \ge r).$$

Of course, the probability is 0 for t < r.

The special case r = 1 (number of trials to the first success) is called the **Geometric** Distribution(p).

The mean of the Geometric Distribution is reasonably straightforward:

$$\mathbf{E} X = \sum_{t=1}^{\infty} t p (1-p)^{t-1} = \frac{1}{p}.$$

(See my notes on sums, paying attention to the fact that I deviously switched the roles of p and 1-p.) Moreover, in Section S1.5 we showed

$$Var X = E(X^2) - (EX)^2 = \frac{1-p}{p^2}.$$

Now observe that a Negative Binomial(r, p) is really the sum of r independent Geometric(p) random variables. (Wait for the first success, start over, wait for the next success, ..., stop after r successes.) Now we use the fact that **expectation is a positive linear operator** to conclude that the mean of Negative Binomial(r, p) is r times the mean the Geometric(p), and the variance of an independent sum is the sum of the variances, so

If $X \sim \text{Negative Binomial}(r, p)$, then

$$\boldsymbol{E} X = \frac{r}{p}$$
 and $\boldsymbol{Var} X = \frac{r(1-p)}{p^2}$

S2.6 Poisson(μ)

A Poisson(μ) random variable N models the count of "successes" when the probability of success is small, but the number of independent trials is large, so that the average success rate is μ .

I will explain the above comment soon. Here is the formal definition of the $\mathrm{Poisson}(\mu)$ distribution

The Poisson(μ) probability mass function is $P(N=k)=e^{-\mu}\frac{\mu^k}{k!},\qquad k=0,1,\ldots$ So $E\,N=\mu,\quad {\rm and}\quad Var\,N=\mu.$





Ladislaus von Bortkiewicz [11] referred to the Poisson distribution as "The Law of Small Numbers." It can be thought of as a peculiar limit of Binomial distributions. Consider a sequence of Binomial(n, p) random variables, where the probability p of success is going to zero, but the number n of trials is growing in such a way that $np = \mu$ remains fixed. Then we have the following.

S2.6.1 Poisson's Limit Theorem For every k,

$$\lim_{n \to \infty} \operatorname{Binomial}(n, \mu/n)(k) = \operatorname{Poisson}(\mu)(k).$$

Proof: Fix n and p, and let $\mu = np$ to rewrite the Binomial probability of k successes in n trials as

$$\frac{n!}{k!(n-k)!}p^{k}(1-p)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!}\frac{(pn)^{k}}{n^{k}}(1-p)^{n}(1-p)^{-k}$$
$$= \frac{\mu^{k}}{k!}\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-k+1}{n}(1-p)^{n}(1-p)^{-k}$$
$$= \frac{\mu^{k}}{k!}\left(1-\frac{\mu}{n}\right)^{n}\frac{n-1}{n}\cdots\frac{n-k+1}{n}\left(1-\frac{\mu}{n}\right)^{-k}$$
$$= \frac{\mu^{k}}{k!}\left(1-\frac{\mu}{n}\right)^{n}\frac{n-1}{n-\mu}\cdots\frac{n-k+1}{n-\mu}.$$

Now consider what happens when we allow n to grow but keep μ fixed by setting $p_n = \mu/n$. It is well known² that $\lim_{n\to\infty} \left(1 - \frac{\mu}{n}\right)^n = e^{-\mu}$, and each of the last k terms has limit 1 as

$$\left(1+\frac{x}{n}\right)^n = \left(e^{\ln\left(1+\frac{x}{n}\right)}\right)^n = e^{n\ln\left(1+\frac{x}{n}\right)}$$

so we first find $\lim_{n\to\infty} n \ln\left(1+\frac{x}{n}\right)$. Note that even if x < 0, for large enough n we will have $1+\frac{x}{n} > 0$, so the logarithm is defined.

² In fact, for any x, we have

 $n \to \infty$. Since k is held fixed, the product of the last k terms converges to 1. So in the limit as $n \to \infty$, the binomial probability of k successes, when the probability of success is μ/n , is just

$$\frac{\mu^{\kappa}}{k!}e^{-\mu}.$$





$$\ln\left(1 + \frac{x}{n}\right) = \ln(1) + \ln'(1)\frac{x}{n} + R(n) = \frac{x}{n} + R(n),$$

$$\lim_{n \to \infty} n \ln \left(1 + \frac{x}{n} \right) = \lim_{n \to \infty} n \left(\frac{x}{n} + R(n) \right) = x + \lim_{n \to \infty} n R(n) = x.$$

Since the exponential function is continuous,

$$\lim_{n \to \infty} \ln(1 + \frac{x}{n})^n = \lim_{n \to \infty} e^{n \ln\left(1 + \frac{x}{n}\right)} = e^{\lim_{n \to \infty} n \ln\left(1 + \frac{x}{n}\right)} = e^x.$$



S2.7 The Normal family

According to the Central Limit Theorem, the limiting distribution of the standardized sum a large number of independent random variables, each with negligible variance relative to the total, is a Normal distribution.

The $N(\mu, \sigma^2)$ density is

$$f_{(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

The mean of an $N(\mu, \sigma^2)$ random variable X

$$\boldsymbol{E} X = \mu$$

and variance

 $Var X = \sigma^2$.

The cdf has no closed form and is simply denoted $\Phi_{(\mu,\sigma^2)}$.

S2.7.1 The Standard Normal

The standard normal has $\mu = 0$ and $\sigma^2 = 1$, so the density is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

It has mean

$$\boldsymbol{E}X=0$$

 $^{^{2}}$ N.B. MATHEMATICA and R parameterize the Normal by its mean and standard deviation, not its variance.

and variance

$$Var X = 1.$$

The cdf has no closed form and is simply denoted Φ .

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx.$$

S2.7.2 The Normal Family

If Z is a standard normal, then

and if

$$X \sim N(\mu, \sigma^2)$$
, then $\frac{x - \mu}{\sigma} \sim N(0, 1)$.

 $\sigma Z + \mu \sim N(\mu, \sigma^2)$

If X and Z are independent normals, then aX + bZ is also normal.









S2.8 Binomial and Normal

S2.8.1 DeMoivre–Laplace Limit Theorem Let X be Binomial(n, p) random variable. Its standardization is $(X - np)/\sqrt{np(1-p)}$. For any real numbers a, b,

$$\lim_{n \to \infty} P\left(a \leqslant \frac{X - np}{\sqrt{np(1 - p)}} \leqslant b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b z^{-z^2/2} \, dx.$$

In practice, this approximation requires that $n \ge \max\{p/(1-p)p, (1-p)/p\}$.



S2.9 Exponential

The Exponential family is used to model **waiting times**, where what we are waiting for is typically called an arrival, or a failure, or death.

Exponential(λ) has density

and cdf

 $F(t) = 1 - e^{-\lambda t}$

 $f(t) = \lambda e^{-\lambda t}$

and ${\bf survival}\ {\bf function}$

$$G(t) = 1 - F(t) = e^{-\lambda t}.$$

It has a constant **hazard rate** f(t)/G(t) equal to λ . The mean of the Exponential(λ) is

$$\boldsymbol{E}T = \frac{1}{\lambda}$$

and the variance is

$$VarT = \frac{1}{\lambda^2}.$$

The Exponential is **memoryless**, that is,

$$P\big(T < t + s \mid T > t\big) = P(T > s).$$





S2.10 Gamma

The $\operatorname{Gamma}(r, \lambda)$ family of distributions is a versatile one.

The distribution of the sum of r independent $\text{Exponential}(\lambda)$ random variables, the distribution of the r^{th} arrival time in a Poisson process with arrival rate λ is the $\text{Gamma}(r, \lambda)$ distribution.

The distribution of the sum of squares of n independent standard normals—the $\chi^2(n)$ distribution is the Gamma(n/2, 1/2) distribution.

The general Gamma (r, λ) distribution $(r > 0, \lambda > 0)$ has density given by

$$f(t) = \frac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t} \qquad (t > 0).$$

The parameter r is the shape parameter, and λ is the scale parameter.

S2.10.1 The Γ function

The Gamma function is defined by

$$\Gamma(r) = \int_0^\infty t^{r-1} e^{-t} \, dt$$

It satisfies $\Gamma(r+1) = r\Gamma(r)$ and $\Gamma(n) = (n-1)!$.

The mean and variance of a $\operatorname{Gamma}(r, \lambda)$ random variable are given by

$$\boldsymbol{E} X = \frac{r}{\lambda}, \qquad \boldsymbol{Var} X = \frac{r}{\lambda^2}.$$

Unfortunately there are two definitions of the Gamma distribution in use! The one I just gave is the one used by Pitman [10], Feller [3], Cramér [2], and Larsen and Marx [9]. The other definition replaces λ by $1/\lambda$ in the density, and is used by Casella and Berger [1], and also by R and MATHEMATICA.





4

2

0.1

8

6



S2.11 Cauchy

If X and Y are independent standard normals, then Y/X has a **Cauchy distribution**. The Cauchy distribution is also the distribution of the tangent of an angle randomly selected from $[-\pi, \pi]$.

If C is a Cauchy random variable, then so is 1/C.

The density is

$$f(x) = \frac{1}{\pi(1+x^2)}$$

and the cdf is

$$F(t) = \frac{1}{\pi} \arctan(t) + \frac{1}{2}.$$

S2.11.1 The expectation of the Cauchy does not exist!

$$\int_{0}^{t} \frac{x}{\pi(1+x^{2})} \, dx = \frac{\ln(1+t^{2})}{2\pi} \xrightarrow[t \to \infty]{} \infty$$
$$\int_{-t}^{0} \frac{x}{\pi(1+x^{2})} \, dx = \frac{-\ln(1+t^{2})}{2\pi} \xrightarrow[t \to \infty]{} -\infty$$

 \mathbf{so}

 $\int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)} \, dx \text{ is divergent and meaningless.}$









The Law of Large Numbers versus the Cauchy Distribution



This depicts a simulation of S_n/n , where S_n is the sum of n simulated Cauchy random variables, for n = 1, ..., 10,000,000,000.

S2.12 Uniform distribution on an interval

Density:
$$f(x) \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise.} \end{cases}$$
$$CDF: \quad F(x) \begin{cases} \frac{x-a}{b-a} & x \in [a,b] \\ 0 & \text{otherwise.} \end{cases}$$
$$E X = \frac{a+b}{2}, \qquad Var X = \frac{(b-a)^2}{12}$$





S2.13 Beta

The $k^{\rm th}$ order statistic of a sample of n independent U[0,1] random variables has a ${\rm Beta}(k,n-k+1)$ distribution.

The density is

$$f(x) = \begin{cases} \frac{1}{\beta(r,s)} x^{r-1} (1-x)^{s-1} & x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

where

$$\beta(r,s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)},$$

where the Gamma function satisfies $\Gamma(n) = (n-1)!$ and $\Gamma(s+1) = s\Gamma(s)$ for s > 0. It has mean and variance given by

$$E X = \frac{r}{r+s}$$
 $Var X = \frac{rs}{(r+s)^2(1+r+s)}.$









S2.14 The $\chi^2(n)$ distribution

Let Z_1, \ldots, Z_n be independent standard normals.

The Chi-square distribution with *n* degrees of freedom, $\chi^2(n)$, is the distribution of $R_n^2 = Z_1^2 + \cdots + Z_n^2.$

One can show (Pitman [10, pp. 358, 365]) that the density of R^2 is given by

$$f_{R_n^2}(t) = \frac{1}{2^{n/2} \Gamma(n/2)} t^{(n/2)-1} e^{-t/2}, \qquad (t > 0),$$

and we write

$$R_n^2 \sim \chi^2(n).$$

This is also the Gamma(n/2, 1/2) distribution.³

The mean and variance are easily characterized:

 $\boldsymbol{E} R_n^2 = n, \qquad \boldsymbol{Var} R_n^2 = 2n.$

It also follows from the definition that if X and Y are independent and

$$X \sim \chi^2(n)$$
 and $Y \sim \chi^2(m)$, then $(X + Y) \sim \chi^2(n + m)$.



³Recall that the Gamma distribution has two naming conventions. In the other convention this would be the Gamma(n/2, 2) distribution.







S2.15 Weibull

The **Weibull distribution** generalizes the exponential distribution as a model of waiting times. It allows for the hazard rate to be age-dependent. There are two parameters of the Weibull distribution, usually denoted λ and α . The density is given by

$$f(t) = \alpha \lambda^{\alpha} t^{\alpha - 1} e^{-(\lambda t)^{\alpha}}$$

and the cumulative distribution function is

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}},$$

so the survival function is

$$G(t) = e^{-(\lambda t)^{\alpha}}$$

and the hazard rate is

$$\lambda(t) = \frac{f(t)}{G(t)} = \alpha \lambda^{\alpha} t^{\alpha - 1}.$$

Note that when $\alpha = 1$, this reduces to the Exponential(λ) distribution. When $\alpha > 1$, the hazard rate is increasing with age, and when $\alpha < 1$, the hazard rate is decreasing with age. See Figure S2.1. N.B. Forbes, Evans, Hastings, and Peacock [4, p. 193] refer to the parameters of the Weibull as η and β , where β is my α and $\eta = 1/\lambda$. MATHEMATICA and Jacod and Protter [5, Example 6, p. 43] use β to mean my $\lambda^{1/\alpha}$. I am using the terminology in Pitman [10, Exercise 4, p. 301].

If T has a Weibull(λ, α) distribution its moments are given by

$$\boldsymbol{E}\,T^n = \Gamma\left(1+\frac{n}{\alpha}\right)\lambda^{-n/\alpha}.$$

S2.16 Logisitc

The Logistic(μ , β) distribution, where $\beta > 0$ has cumulative distribution function given by

$$F(x) = \frac{1}{1 + e^{-(x-\mu)/\beta}}$$

and density

$$f(x) = \frac{1}{\left(1 + e^{-(x-\mu)/\beta}\right)^2 \beta} e^{-(x-\mu)/\beta}$$

It has mean and median μ and β is a scale parameter. The variance is $\pi^2 \beta^2/3$.

The logistic density looks a lot like the normal density, see Figure S2.2.

Bibliography

- G. Casella and R. L. Berger. 2002. Statistical inference, 2d. ed. Belmont, California: Brooks/Cole Cengage Learning.
- [2] H. Cramér. 1946. Mathematical methods of statistics. Number 34 in Princeton Mathematical Series. Princeton, New Jersey: Princeton University Press. Reprinted 1974.
- [3] W. Feller. 1968. An introduction to probability theory and its applications, 3d. ed., volume 1. New York: Wiley.
- [4] C. Forbes, M. Evans, N. Hastings, and B. Peacock. 2011. Statistical distributions, 4th. ed. Hoboken, New Jersey: John Wiley & Sons.
- [5] J. Jacod and P. Protter. 2004. Probability essentials, 2d. ed. Berlin and Heidelberg: Springer.
- [6] N. L. Johnson and S. Kotz. 1969. Discrete distributions. Boston: Houghton Mifflin.
- [7] . 1970. Continuous univariate distributions I. New York: Houghton Mifflin.
- [8] . 1970. Continuous univariate distributions II. Boston: Houghton Mifflin.

v. 2020.03.26::15.07







Figure S2.1.





Figure S2.2.

- [9] R. J. Larsen and M. L. Marx. 2012. An introduction to mathematical statistics and its applications, fifth ed. Boston: Prentice Hall.
- [10] J. Pitman. 1993. Probability. Springer Texts in Statistics. New York, Berlin, and Heidelberg: Springer.
- [11] L. von Bortkiewicz. 1898. Das Gesetz der kleinen Zahlen [The law of small numbers]. Leipzig: B.G. Teubner. The imprint lists the author as Dr. L. von Bortkewitsch.