Distributions

Binomial

If X is the number of successes in n independent Bernoulli trials with probability of success p then $X \stackrel{d}{=} \mathrm{Bi}(n, p)$.

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$\mathbb{E}(X) = np$$
$$\mathrm{Var}(X) = np(1-p)$$

X has recursive formula:

$$\frac{p_X(x)}{p_X(x-1)} = \frac{\frac{n+1}{x} - 1}{\frac{1}{p} - 1}$$

Geometric

If N is the number of failures before a success in a sequence of independent Bernoulli trials with probability of success p then $N \stackrel{d}{=} \mathrm{G}(p)$.

$$p_N(n) = (1-p)^n p$$
$$\mathbb{E}(N) = \frac{1-p}{p}$$
$$\mathrm{Var}(N) = \frac{1-p}{p^2}$$

The geometric distribution has the memoryless property.

Negative Binomial

If Z is the number of failures before the r^{th} success in a sequence of independent Bernoulli trials with probability of success p then $Z \stackrel{d}{=} \mathrm{Nb}(r, p)$.

$$p_Z(z) = \binom{z+r-1}{r-1} p^r (1-p)^z$$
$$= \binom{-r}{z} p^r (p-1)^z$$

Where $\binom{x}{k} = \frac{x(x-1)(x-2)\ldots(x-k+2)(x-k+1)}{k!}$ for $x \in \mathbb{R}$

Hypergeometric

When sampling n items from a population of N without replacement, where D are defective, the amount of defective items selected is X and; $X \stackrel{d}{=} \mathrm{Hg}(n, D, N)$.

$$p_X(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}$$
$$\mathbb{E}(X) = \frac{nD}{N}$$
$$\mathrm{Var}(X) = \frac{nD(N-D)}{N^2} \cdot (1 - \frac{n-1}{N-1})$$

Poisson

An analogue for the Binomial distribution in continuous time. If α is the expected amount of successes over unit time (the Poisson rate), then the amount of successes $X \stackrel{d}{=} \mathrm{Pn}(\alpha)$ and the amount of successes in t units of time $\stackrel{d}{=} \mathrm{Pn}(\alpha t)$.

$$p_X(x) = \frac{e^{-\alpha}(\alpha)^x}{x!}$$
$$\mathbb{E}(X) = \mathrm{Var}(X) = \alpha$$

The Poisson distribution has the memoryless property and can be used to approximate the Binomial distribution for small p, ($p < 0.05$) where $\mathrm{Bi}(n, p) \stackrel{d}{\approx} \mathrm{Pn}(np)$.

Discrete Uniform

A representation of discrete events with equal probabilities of all outcomes, for an X that can take integer values between m and n, $X \stackrel{d}{=} \mathrm{U}(m, n)$.

$$p_X(x) = \frac{1}{n-m+1}$$
$$\mathbb{E}(X) = \frac{m+n}{2}$$
$$\mathrm{Var}(X) = \frac{1}{12}((n-m+1)^2 - 1)$$

Continuous Uniform

A representation of continuous events with equal probabilities, for an X that can take any real value between a and b, $X \stackrel{d}{=} \mathrm{R}(a, b)$.

$$f_X(x) = \frac{1}{b-a}$$
$$\mathbb{E}(X) = \frac{a+b}{2}$$
$$\mathrm{Var}(X) = \frac{1}{12}(b-a)^2$$

Exponential

A continuous analogue of the geometric distribution, modelling the waiting time until an event occurs in continuous time. For an event that has probability of α of occurring over one period of time, the waiting time until the event occurs T follows $T \stackrel{d}{=} \exp(\alpha)$.

$$f_T(t) = \begin{cases} \alpha e^{-\alpha t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$
$$\mathbb{E}(T) = \frac{1}{\alpha}$$
$$\mathrm{Var}(T) = \frac{1}{\alpha^2}$$

The exponential distribution has the lack of memory property such that $\mathbb{P}(T \geq x + y \mid T \geq x) = \mathbb{P}(T \geq y)$.

Gamma

A continuous analogue of the negative binomial distribution, modelling the waiting time until r events occur in continuous time. For an event that has probability of α of occurring over one period of time, the waiting time until the r^{th} event occurs - T - follows $T \stackrel{d}{=} \gamma(r, \alpha)$.

$$f_T(t) = \frac{\alpha^r t^{r-1} e^{-\alpha t}}{\Gamma(r)},\ t > 0$$
$$F_T(t) = 1 - \sum_{k=0}^{r-1} \frac{(\alpha t)^k}{k!} e^{-\alpha t}$$
$$\mathbb{E}(T^k) = \frac{\Gamma(r+k)}{\Gamma(r)\alpha^k}$$
$$\mathbb{E}(T) = \frac{r}{\alpha}$$
$$\mathrm{Var}(T) = \frac{r}{\alpha^2}$$

$$\text{where } \Gamma(r) = \int_0^\infty e^{-x} x^{r-1} dx = (r-1)\Gamma(r-1) \text{ for all } r > 0$$
$$\Gamma(1) = 1 \text{ and } \Gamma(k) = (k-1)! \text{ for any } k \in \mathbb{N}$$

Beta

A random variable X has a beta distribution with parameters $\alpha > 0$ and $\beta > 0$ and is denoted $X \stackrel{d}{=} \mathrm{Beta}(\alpha, \beta)$ if it has the following properties:

$$f_X(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases}$$
$$\mathbb{E}(X^k) = \frac{\Gamma(\alpha+k)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+k)}$$
$$\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$$
$$\mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$\text{where } B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx.$$

The distribution is $\cup$ shaped for $\alpha = \beta < 1$ and $\cap$ shaped for $\alpha = \beta > 1$ and $\mathrm{Beta}(1,1) = \mathrm{R}(0,1)$.

Pareto

The Pareto distribution is shaped similarly to the exponential distribution, though the tail of an exponential distribution is thinner. A random variable X has a pareto distribution denoted $X \stackrel{d}{=} \mathrm{Pareto}(\alpha, \gamma)$ with parameters $\{\alpha, \gamma\} \in \mathbb{R}^+$ if it has the properties:

$$f_X(x) = \frac{\gamma\alpha^\gamma}{x^{\gamma+1}},\ \text{for } \alpha \le x < \infty$$
$$\mathbb{E}(X) = \frac{\gamma\alpha}{\gamma-1},\ \text{for } \gamma > 1$$
$$\mathrm{Var}(X) = \frac{\gamma\alpha^2}{(\gamma-1)^2(\gamma-2)},\ \text{for } \gamma > 2$$

Normal

The Normal distribution is parameterised by its mean and variance μ and σ^2. A normally distributed random variable X is written $X \stackrel{d}{=} \mathrm{N}(\mu, \sigma^2)$. The case N(0,1) is the *Standard Normal Distribution* and it's pdf is denoted $\varphi(z)$.

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2},\ \text{for } x \in \mathbb{R}$$
$$\mathbb{E}(X) = \mu$$
$$\mathrm{Var}(X) = \sigma^2$$

For $Z \stackrel{d}{=} \mathrm{N}(0,1)$ (the Standard Normal Distribution):

$$\mathbb{E}(Z^{2k}) = \frac{(2k)!}{2^k k!}$$
$$\mathbb{E}(Z^{2k+1}) = 0$$

The Normal distribution can be used to approximate:

- $X \stackrel{d}{=} \mathrm{Bi}(n,p) \stackrel{d}{\approx} \mathrm{N}(np, np(1-p))$ for large n and p not close to 0 or 1. ($np > 5$ and $n(1-p) > 5$)
- $X \stackrel{d}{=} \mathrm{Pn}(\lambda) \stackrel{d}{\approx} \mathrm{N}(\lambda, \lambda)$ for large λ.
- $X \stackrel{d}{=} \gamma(r, \alpha) \stackrel{d}{\approx} \mathrm{N}(\frac{r}{\alpha}, \frac{r}{\alpha^2})$ for large r.

Weibull

The Weibull distribution is used to model survival data, where the hazard function, or rate of failures at a specific time may increase, decrease or be constant. It is parameterised by β and γ, written $X \stackrel{d}{=} \mathrm{Weibull}(\beta, \gamma)$ and has the properties:

$$f_X(x) = \frac{\gamma x^{\gamma-1}}{\beta^\gamma} e^{-(\frac{x}{\beta})^\gamma},\ \text{for } 0 \le x < \infty$$
$$\mathbb{E}(X) = \beta\Gamma\left(\frac{\gamma+1}{\gamma}\right)$$
$$\mathrm{Var}(X) = \beta^2\left[\Gamma\left(\frac{\gamma+2}{\gamma}\right) - \Gamma\left(\frac{\gamma+1}{\gamma}\right)^2\right]$$

The hazard rate of any random variable T is given by $\frac{f_T(t)}{1-F_T(t)}$, giving the Weibull distribution a hazard rate of $\gamma\frac{t^{\gamma-1}}{\beta\gamma}$.

Cauchy

The Cauchy distribution has parameters m - the location, and a - the scale parameter, and is written $X \stackrel{d}{=} \mathrm{C}(m, a)$.

$$f_X(x) = \frac{1}{\pi}\frac{a}{a^2 + (x-m)^2},\ \text{for } -\infty < x < \infty$$

It does not have a defined mean or variance.

Lognormal

If $X \stackrel{d}{=} \mathrm{N}(\mu, \sigma^2)$ and $Y \stackrel{d}{=} e^X$, then Y has a lognormal distribution with the same parameters $Y \stackrel{d}{=} \mathrm{LN}(\mu, \sigma^2)$.

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma y} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}},\ \text{for } y > 0$$
$$\mathbb{E}(Y^r) = e^{r\mu + \frac{1}{2}r^2\sigma^2},\ r \ge 0$$
$$\mathrm{Var}(Y) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$$

Bivariate Normal

A bivariate normal distribution is defined by the means and variances of the univariate normal distributions X and Y and their correlation coefficient ρ. We write $(X,Y) \stackrel{d}{=} \mathrm{N}_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, or for the standard bivariate normal with $\mu_X = \mu_Y = 0$ and $\sigma_X^2 = \sigma_Y^2 = 1$, $\mathrm{N}_2(\rho)$.

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

$$(X|Y=y) \stackrel{d}{=} \mathrm{N}(\mu_X + \rho\sigma_X \frac{(y-\mu_Y)}{\sigma_Y}, \sigma_X^2(1-\rho^2))$$

The correlation coefficient ρ is calculated as $\frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$ where $\mathrm{Cov}(X,Y) = \mathbb{E}((X-\mu_X)(Y-\mu_Y)) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ $= \mathbb{E}(X\mathbb{E}(Y|X)) - \mathbb{E}(X)\mathbb{E}(Y)$.

Convolutions & Linear Combinations

If X and Y are **independent** their sum (convolution) is:

$$\begin{aligned} p_{X+Y}(z) &= \sum_{x \in S_X} p_X(x) p_Y(z-x) \\ &= \sum_{y \in S_Y} p_X(z-y)p_Y(y),\ (X,Y) \text{ discrete,} \\ f_{X+Y}(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) \\ &= \int_{-\infty}^{\infty} f_X(z-y) f_Y(y),\ (X,Y) \text{ continuous.} \end{aligned}$$

And for $Z_1 = aX + bY$, $Z_2 = cX + dY$:

$$\begin{aligned} &\mathbb{E}(Z_1) = a\mathbb{E}(X) + b\mathbb{E}(Y) \\ &\mathrm{Cov}(Z_1, Z_2) = ac\mathrm{Var}(X) + 2ab\mathrm{Cov}(X,Y) + bd\mathrm{Var}(Y) \\ &\mathrm{Var}(Z_1) = a^2\mathrm{Var}(X) + 2ab\mathrm{Cov}(X,Y) + b^2\mathrm{Var}(Y) \end{aligned}$$

where the variance property is a special case; $\mathrm{Cov}(Z_1, Z_1) = \mathrm{Var}(Z_1)$.

Conditioning on Random Variables

$\mathbb{E}(X|Y) = \mathbb{E}(X|Y=y)$ is the *conditional expectation* of X given Y.

$\mathbb{E}(X|Y) := \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$, replacing the integral for a sum over $x \in S_X$ for the discrete case.

The **Law of Total Expectation** is $\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$, where we take the outer integral over all values of y and then rearrange from $\int_{-\infty}^{\infty}(\int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx) f_Y(y)dy$ to get $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{(X,Y)}(x,y)dydx = \int_{-\infty}^{\infty} x f_X(x)dx$.

Then for variance, $\mathrm{Var}(X) = \mathrm{Var}(\mathbb{E}(X|Y)) + \mathbb{E}(\mathrm{Var}(X|Y))$.

Generating Functions

Probability Generating Functions (PGF's)

The PGF exists only for non-negative integer valued random variables and the PGF of X is given by

$$P_X(z) = \mathbb{E}(z^X) = \sum_{x=0}^{\infty} p_X(x) z^x.$$

We then get that

$$p_X(k) = \mathbb{P}(X=k) = \frac{P_X^{(k)}(0)}{k!}$$

Properties of the PGF

- $P_X(z) = \mathbb{E}(z^X) = \mathbb{E}(\mathbb{E}(z^X|Y))$.
- $P_X'(1) = \mathbb{E}(X)$ and $P_X''(1) = \mathbb{E}(X(X-1))$,
- $\implies \mathrm{Var}(X) = P_X''(1) + P_X'(1) - P_X'(1)^2$.
- For independent X and Y, $P_{X+Y}(z) = P_X(z)P_Y(z)$ since $\mathbb{E}(z^{X+Y}) = \mathbb{E}(z^X)\mathbb{E}(z^Y)$.

Binomial PGF If $X \stackrel{d}{=} \mathrm{Bi}(n,p)$, then:

$$P_X(z) = (1-p+pz)^n, \text{ for } z \in \mathbb{R}.$$

Poisson PGF If $X \stackrel{d}{=} \mathrm{Pn}(\lambda)$, then:

$$P_X(z) = e^{-\lambda(1-z)}, \text{ for } z \in \mathbb{R}.$$

Negative Binomial PGF If $X \stackrel{d}{=} \mathrm{Nb}(r,p)$, then:

$$P_X(z) = p^r(1-(1-p)z)^{-r}, \text{ for } |z| < \frac{1}{1-p}.$$

The PGF for the **geometric** distribution can be derived from this as $X \stackrel{d}{=} \mathrm{G}(p) \stackrel{d}{=} \mathrm{Nb}(1,p)$

Moment Generating Functions (MGF's)

The k^{th} moment (about the origin) of a random variable X is $\mu_k = \mathbb{E}(X^k)$.

The k^{th} *central* moment (about the mean) of a random variable X is $\nu_k = \mathbb{E}((X-\mu)^k)$.

The MGF is defined over all $t \in \{t : \mathbb{E}(e^{tX}) < \infty\}$ as

$$M_X(t) = \mathbb{E}(e^{tX}) = \int_{x \in S_X} e^{tx} f_X(x)dx,$$

replacing the integral with a sum in the discrete case as the MGF is defined for both.

Properties of the MGF

- $M_X(0) = 1$, $M_X'(0) = \mathbb{E}(X)$ and $M_X''(0) = \mathbb{E}(X^2)$,
- $\implies \mathrm{Var}(X) = M_X''(0) - M_X'(0)^2$.
- $\mu_k = M_X^{(k)}(0)$.
- If $Y = aX + b$ then $M_Y(t) = e^{bt}M_X(at)$.
- For **independent** X and Y, $M_{X+Y}(t) = M_X(t)M_Y(t)$.
- If X is a discrete RV defined on the non-negative integers then $M_X(t) = P_X(e^t)$ and $P_X(z) = M_X(\log z)$.
- The central moment generating function $N_X(t) = \mathbb{E}\left(e^{(X-\mu)t}\right)$

Binomial MGF If $X \stackrel{d}{=} \text{Bi}(n, p)$, then:

$$M_X(t) = ((1-p) + pe^t)^n$$

Geometric MGF If $X \stackrel{d}{=} \text{G}(p)$, then:

$$M_X(t) = \frac{pe^t}{1-(1-p)e^t}$$

Exponential MGF If $X \stackrel{d}{=} \exp(\alpha)$, then:

$$M_X(t) = \frac{\alpha}{\alpha - t}$$

Poisson MGF If $X \stackrel{d}{=} \text{Pn}(\lambda)$, then:

$$M_X(t) = e^{\lambda(e^t-1)} = 1 + \lambda t + \lambda(\lambda+1)\frac{t^2}{2} + \dots$$

Gamma MGF If $X \stackrel{d}{=} \gamma(r, \alpha)$, then:

$$M_X(t) = \left(1 - \frac{t}{\alpha}\right)^{-r} = 1 + \frac{rt}{\alpha} + \frac{r(r+1)}{\alpha^2}\frac{t^2}{2} + \dots$$

Normal MGF If $X \stackrel{d}{=} \text{N}(\mu, \sigma^2)$, then:

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}, \text{ and}$$
$$N_X(t) = e^{\frac{1}{2}\sigma^2 t^2}$$

Cumulant Generating Functions (CGF's)

The CGF of a random variable X is given by

$$K_X(t) = \ln M_X(t)$$

and $\kappa_r = K_X^{(r)}(0)$ is the r^{th} cumulant of X.

Properties of the CGF

- For **independent** X and Y, $K_{X+Y}(t) = K_X(t) + K_Y(t)$.
- $\kappa_1 = \mathbb{E}(X)$.
- $\kappa_2 = \text{Var}(X)$.
- $\kappa_3 = \mathbb{E}(X^3)$ is the *skewness* of X.
- $\kappa_4 = \mathbb{E}(X^4) - 3\sigma^4$ is the *kurtosis* of X.

The **coefficient of skewness** skew(X) is then $\frac{\kappa_3}{\sigma^3}$ and **coefficient of kurtosis** kurt(X) is $\frac{\kappa_4}{\sigma^4}$.

Other Formulae

Expected Values via. Tail Probabilities

For a positive RV X, $(\mathbb{P}(X \geq 0) = 1)$

$$\mathbb{E}(X^n) = n\int_0^\infty x^{n-1}(1 - F_X(x))dx$$

Chebyshev's Inequality

$$\mathbb{P}\left(\frac{|X-\mu|}{\sigma} \geq k\right) \leq \frac{1}{k^2}.$$

Common Series

$$\sum_{k=0}^\infty \frac{x^k}{k!} = e^x \qquad \text{(Taylor Series)}$$
$$\sum_{i=1}^n ar^{i-1} = \frac{a(1-r^n)}{1-r}, \qquad r \neq 1$$
$$\sum_{k=0}^\infty ar^k = \frac{a}{1-r}, \qquad r \in [0,1)$$
$$\sum_{n=1}^\infty \frac{1}{n^p} \quad \text{diverges for } p \leq 1$$

Central Limit Theorem

If $X_1, X_2, \dots$ are independent identically distributed random variables with $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$ and $S_n = X_1 + X_2 + \dots + X_n$ then;

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \text{N}(0,1) \text{ as } n \to \infty.$$

Put otherwise, as $n \to \infty$, $S_n \xrightarrow{d} \text{N}(n\mu, n\sigma^2)$ or the sample mean $\overline{X} \xrightarrow{d} \text{N}(\mu, \frac{\sigma^2}{n})$.

Laplace Transform

The Laplace transform of a RV X is defined as

$$L_X(t) = M_X(-t) = \mathbb{E}(e^{-tX}).$$

Then the inversion formula

$$F_X(x) = \lim_{t\to\infty} \sum_{k \leq tx} \frac{(-t)^k}{k!} L_X^{(k)}(t)$$

can be used to recover the cumulative distribution of X.

Axioms of the Probability Function

1. $\mathbb{P}(A) \geq 0$
2. $\mathbb{P}(\Omega) = 1$
3. For *disjoint* sets A_i, $\mathbb{P}(\bigcup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mathbb{P}(A_i)$.

Solving *min*/*max* Questions

For for independent U and V and $Z = min(U, V)$:

$$\begin{aligned}\mathbb{P}(Z \leq z) &= 1 - \mathbb{P}(Z > z) = 1 - \mathbb{P}(U > z, V > z) \\ &= 1 - \mathbb{P}(U > z)\mathbb{P}(V > z)\end{aligned}$$

$$Z = max(U, V) \implies \mathbb{P}(Z \leq z) = \mathbb{P}(U \leq z, V \leq z)$$

Conditional Variance

$$\text{Var}(X|Y=y) = E(X^2|Y=y) - E(X^2|Y=y)^2.$$