Discrete Distributions

Bernoulli (p)

Models the probability of an event succeeding or failing, with $x = 1$ as a success and $x = 0$ as fail. There is only one trial.

Given $x = 0, 1; \quad 0 \leq p \leq 1$

- p is the probability of getting selected trait
- Has p probability of being 1 and $1 - p$ probability of being 0

$P(X = x) = p^x(1-p)^{1-x}$
$\text{CDF} = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$
$\mu = p$
$\sigma^2 = p(1-p)$
$M(t) = (1-p) + pe^t$

Binomial (n, p)

Models the probability of getting x successes out of n trials with p probability of each trial succeeding.

Given $x = 0, 1, 2, \ldots, n; \quad 0 \leq p \leq 1$

- p is probability of selecting a particular trait
- n is number of samples in a round of sampling
- Predicts probability of getting certain number of chosen trait in sample set

$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$
$\mu = np$
$\sigma^2 = np(1-p)$
$M(t) = (pe^t + (1-p))^n$

Discrete Uniform (N)

Models the chance of x as a result when all numbers from 1 to N are equally likely.

Given $x = 1, 2, \ldots, N; \quad N = 1, 2, \ldots$

- N is the largest possible sample
- All numbers from 1 to N are equally likely

$P(X = x) = 1/N$
$\text{CDF} = \frac{\lfloor x \rfloor + 1}{N}$
$\mu = \frac{N+1}{2}$
$\sigma^2 = \frac{(N+1)(N-1)}{12}$
$M(t) = \frac{1}{N} \sum_{i=1}^{N} e^{it}$

Geometric (p)

Models chance of taking x trials to get a success, with each success having p chance of succeeding.

Given $x = 1, 2, \ldots; \quad 0 \leq p \leq 1$

- p is probability of getting certain trait
- Predicts number of samples needed to get a sample of particular trait

$P(X = x) = p(1-p)^{x-1}$
$\text{CDF} = \begin{cases} 1 - (1-p)^{\lfloor x \rfloor} & x \geq 0 \\ 0 & x < 0 \end{cases}$
$\mu = 1/p$
$\sigma^2 = \frac{1-p}{p^2}$
$M(t) = \frac{pe^t}{1-(1-p)e^t}$

Hypergeometric (N, K, M)

Predicts likelihood of selecting x successes after K trials of selecting from a population of N with M samples within that are successes without replacement.

Given $x = 0, 1, 2, \ldots, K; \quad M - (N - K) \leq x \leq M; \quad N, M, K = 0, 1, 2, \ldots$

- N is the population size
- M is the number of samples in the population with a certain trait
- K number of samples taken in a round of sampling
- Predicts the likelihood of selecting X samples of type M after selecting K samples from population N

$P(X = x) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}$
$\mu = KM/N$
$\sigma^2 = \frac{KM(N-M)(N-K)}{N^2(N-1)}$

Negative Binomial (r, p)

Predicts likelihood of sampling r successes of p probability after $x + r$ samples

Given $x = 0, 1, 2, \ldots; \quad 0 \leq p \leq 1$

- p is the probability of getting a particular trait in one sample
- r is the desired number of samples with a particular trait
- Predicts number of likelihood of getting r samples of trait after $X + r$ samples

$P(X = x) = \binom{r + x - 1}{x} p^r (1-p)^x$
$\mu = \frac{r(1-p)}{p}$
$\sigma^2 = \frac{r(1-p)}{p^2}$
$M(t) = (\frac{p}{1-(1-p)e^t})^r$

Poisson Distribution (λ)

Assumes chances of an event happening in a short time is proportional to a large time. Predicts likelihood of x events happening in the next unit if λ events happen on average per unit where the unit does not need to be time.

- $N_0 = 0$
- For $s > 0$, N_t and $N_{t+s} - N_t$ are independent random variables
- N_s and $N_{t+s} - N_t$ are the same distribution
- $\lim_{t \downarrow 0} \frac{P(N_t = 1)}{t} = \lambda$ (Average rate of events will be λ everywhere)
- $\lim_{t \downarrow 0} \frac{P(N_t > 1)}{t} = 0$ (Can't have two events at the same time)

Given $x = 0, 1, 2, \ldots; \quad 0 \leq \lambda$

- λ is the number of times on average an event will happen within an interval
- Predicts number of times an event will happen within an interval
- Approximates the Binomial Distribution

$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$
$\mu = \lambda$
$\sigma^2 = \lambda$
$M(t) = e^{\lambda(e^t - 1)}$

Continuous Distributions

Beta (α, β)

Often used to model proportions that lie on $(1, 0)$

Given $0 \leq x \leq 1; \quad \alpha > 0; \quad \beta > 0$

$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$
$\mu = \frac{\alpha}{\alpha+\beta}$
$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
$M(t) = 1 + \sum_{k=1}^{\infty} \frac{t^k}{k!} \prod_{r=0}^{k-1} \frac{a+r}{\alpha+\beta+r}$

Cauchy (θ, σ)

Symmetric, bell shaped, and has no mean, often used as an extreme edge case for statistical theory, but also represents the ratio between two normal distributions among other unsuspecting things.

Given $-\infty < x < \infty; \quad -\infty < \theta < \infty; \quad \sigma > 0$

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

Chi squared (p)

Given $0 \leq x < \infty; \quad p = 1, 2, 3, \ldots$

$f(x) = \frac{x^{p/2-1}e^{-x/2}}{\Gamma(p/2)2^{p/2}}$
$\mu = p$
$\sigma^2 = 2p$
$M(t) = (\frac{1}{1-2t})^{p/2}$

$N(0,1)^2 = \text{Chi}(1)$

Double Exponential (μ, σ)

Given by reflecting the exponential around its mean. Does not have a bell shape.

Given $-\infty < x < \infty; \quad -\infty < \mu < \infty; \quad \sigma > 0$

$f(x) = \frac{e^{-|x-\mu|/\sigma}}{2\sigma}$
$\mu = \mu$
$\sigma^2 = 2\sigma^2$
$M(t) = \frac{e^{\mu t}}{1-(\sigma t)^2}$

Exponential (β)

Given $0 \leq x < \infty; \quad \beta > 0$

$f(x) = \frac{e^{-x/\beta}}{\beta}$
$F(x) = 1 - e^{-x/\beta}$
$\mu = \beta$
$\sigma^2 = \beta^2$
$M(t) = \frac{1}{1-\beta t}$

F (v_1, v_2)

Given $0 \leq \infty; \quad v_1, v_2 = 1, 2, 3, \ldots$

$f(x) = \frac{\Gamma(\frac{v_1+v_2}{2})}{\Gamma(\frac{v_1}{2})\Gamma(\frac{v_2}{2})} (\frac{v_1}{v_2})^{v_1/2} \frac{x^{(v_1-2)/2}}{(1+\frac{v_1 x}{v_2})^{(v_1+v_2)/2}}$
$\mu = \frac{v_2}{v_2-2}$
$\sigma^2 = 2(\frac{v_2}{v_2-2})^2 \frac{v_1+v_2-2}{v_1(v_2-4)}$
$EX^n = \frac{\Gamma(\frac{v_1+2n}{2})\Gamma(\frac{v_2-2n}{2})}{\Gamma(v_1/2)\Gamma(v_2/2)} (\frac{v_2}{v_1})^n \quad ; n < \frac{v_2}{2}$

$\text{F}(a,b) = \frac{\frac{\chi_a^2}{a}}{\frac{\chi_b^2}{b}}$

Gamma (α, β)

α being the shape parameter and β being the scale parameter, it is also a generalization of the exponential, erlang, weibull, and chi-squared distrbutions.

Given $0 \leq x < \infty; \quad \alpha, \beta > 0$

$f(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}$
$\mu = \alpha\beta$
$\sigma^2 = \alpha\beta^2$
$M(t) = (\frac{1}{1-\beta t})^\alpha$

Logistic (μ, β)

Given $-\infty < x < \infty; \quad -\infty < \mu < \infty; \quad \beta > 0$

$f(x) = \frac{e^{-(x-\mu)/\beta}}{\beta(1+e^{-(x-\mu)/\beta})^2}$
$\mu = \mu$
$\sigma^2 = \frac{\pi^2\beta^2}{3}$
$M(t) = e^{\mu t}\Gamma(1+\beta t)$

Lognormal (μ, σ^2)

Distribution where the log of a variable is normally distributed. Similar in appearance to the gamma distribution.

Given $0 \leq x < \infty; \quad -\infty < \mu < \infty; \quad \sigma > 0$

$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\frac{e^{-(\log x-\mu)^2/(2\sigma^2)}}{x}$
$\mu = e^{\mu+(\sigma^2/2)}$
$\sigma^2 = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$
$EX^n = e^{n\mu+n^2\sigma^2/2}$

Normal (μ, σ^2)

Can be used to approximate many different kinds of distributions as their population sizes increase.

Given $-\infty < x < \infty; \quad -\infty < \mu < \infty; \quad \sigma > 0$

$f(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma}$
$\mu = \mu$
$\sigma^2 = \sigma^2$
$M(t) = e^{\mu t+\sigma^2t^2/2}$

Paretto (α, β)

Given $a < x < \infty; \quad \alpha, \beta > 0$

$f(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}}$
$\mu = \frac{\beta\alpha}{\beta-1} \quad ;\beta > 1$
$\sigma^2 = \frac{\beta\alpha^2}{(\beta-1)^2(\beta-2)} \quad ;\beta > 2$

T (v)

Given $-\infty < x < \infty; \quad v = 1, 2, 3, \ldots$

$f(x) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})}\frac{1}{\sqrt{v\pi}}\frac{1}{(1+(\frac{x^2}{v}))^{(v+1)/2}}$
$\mu = 0 \quad ;v > 1$
$\sigma^2 = \frac{v}{v-2} \quad ;v > 2$
$MX^n = \begin{cases} \frac{\Gamma(\frac{n+1}{2})\Gamma(\frac{v-n}{2})}{\sqrt{\pi}\Gamma(v/2)}v^{n/2} & n < v; n \text{ is even} \\ 0 & n < v; n \text{ is odd} \end{cases}$

$\mathrm{T}(v) = \frac{N(0,1)}{\sqrt{\frac{\chi^2_v}{v}}}$

Uniform (a, b)

All values between a and b are evenly distributed and x has equal chance of landing anywhere on that range.

Given $a \leq x \leq b$

- a is the lower bound of the distribution
- b is the upper bound
- All values between a and b are equally distributed

$f(x) = \frac{1}{b-a}$
$\mu = \frac{b+a}{2}$
$\sigma^2 = \frac{(b-a)^2}{12}$
$M(t) = \frac{e^{bt}-e^{at}}{t(b-a)}$

Weibull (γ, β)

Given $0 \leq x < \infty; \quad \gamma, \beta > 0$

$f(x) = \frac{\gamma}{\beta}x^{\gamma-1}e^{-x^\gamma/\beta}$
$\mu = \beta^{1/\gamma}\Gamma(1+\frac{1}{\gamma})$
$\sigma^2 = \beta^{2/\gamma}(\Gamma(1+\frac{2}{\gamma}) - \Gamma^2(1+\frac{1}{\gamma}))$
$EX^n = \beta^{n/\gamma}\Gamma(1+\frac{n}{\gamma})$

Multivariable Distributions

Covariance and Correlation

$\mathrm{Cov}(X,Y) = E((X-\mu_X)(Y-\mu_Y))$
$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$

Multinomial Distribution

Given $\vec{\mu} \in \Re^k; \quad \mathbf{\Sigma} \in \Re^{k^2}; \quad k \in \mathbb{N}$

$f(\vec{x}) = (2\pi)^{-k/2}\det(\mathbf{\Sigma})^{-1/2}\exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T\mathbf{\Sigma}^{-1}(\vec{x}-\vec{\mu})\right)$
$\vec{\mu} = \vec{\mu}$
$\boldsymbol{\sigma^2} = \mathbf{\Sigma}$
$M(\vec{t}) = \exp\left(\vec{\mu}^T\vec{t} + \frac{1}{2}\vec{t}^T\mathbf{\Sigma}\vec{t}\right)$

Bivariate case $(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$

$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{\vec{x}-\vec{\mu}_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{\vec{x}-\vec{\mu}_x}{\sigma_X}\right)\left(\frac{\vec{y}-\vec{\mu}_Y}{\sigma_Y}\right) + \left(\frac{\vec{y}-\vec{\mu}_Y}{\sigma_Y}\right)^2\right)\right)$

$\vec{\mu} = \langle\mu_X, \mu_Y\rangle$
$\Sigma = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$

Families of Distributions

Exponential Family

Exponential Family
Any statistical distribution or family of distributions that can fit into the form:
$f(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^k w_i(\theta)t_i(x)\right)$

Binomial

$\binom{n}{x}(1-p)^n\exp\left(\log(\frac{p}{1-p})x\right)$

Normal

$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\mu^2}{2\sigma^2}\right)\exp\left(-\frac{x^2}{2\sigma^2}+\frac{\mu x}{\sigma^2}\right)$

Location Scale Family

Location Scale Family
Any statistical distribution or family of distributions that can fit into the form:
$g(x|\mu,\sigma) = \frac{1}{\sigma}f(\frac{x-\mu}{\sigma})$

Probability Inequalities

Chebychev's Inequality

Let X be a random variable and let $g(x)$ be a non-negative function. Then, for any $r > 0$

$P(g(X) \geq r) \leq \frac{Eg(X)}{r}$

Normal Probability Inequality

With Z as a normal distribution,

Very similar to a binomial, except there is more than one possible outcome per trial — as compared to success or failure in the binomial.

- Taking the marginal of any of the possible outcomes results in a regular binomial

Multivariable Normal $(\vec{\mu}, \mathbf{\Sigma})$

$P(|Z| \geq t) \leq \sqrt{\frac{e}{\pi}}\frac{e^{-t^2/2}}{t}$ for all $t > 0$

Random Samples

Properties of the sample

Mean: $\bar{X} = \frac{X_1+\ldots+X_n}{n} = \frac{1}{n}\sum_{i=1}^n X_i$

Variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2$

$E(\sum_{i=1}^n g(X_i)) = nEg(X_1)$
$\mathrm{Var}(\sum_{i=1}^n g(X_i)) = n\mathrm{Var}g(X_1)$

Properties of properties of the sample of random variables

$E(\bar{X}) = \mu$
$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$
$E(S^2) = \sigma^2$

Sample distributions of common distributions

Normal (μ, σ^2)

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
S^2 is independent from $\bar{X}$

Gamma (α, β)

$\bar{X} \sim \mathrm{Gamma}\left(n\alpha, \frac{\beta}{n}\right)$

Cauchy

$\mathrm{Cauchy}(0,\sigma_1)+\ldots+\mathrm{Cauchy}(0,\sigma_n) = \mathrm{Cauchy}(0, \sum_{i=1}^n \sigma_i)$

Chi-squared

$\mathrm{Chi}(a_1)+\ldots+\mathrm{Chi}(a_n) = \mathrm{Chi}(\sum_{i=1}^n a_i)$