Probability Distribution Properties

1 Discrete Distributions

1.1 Uniform Random Variable

$$E[X] = \frac{x_1 + \cdots + x_n}{n}; \quad Var(X) = \frac{x_1^2 + \cdots + x_n^2}{n} - \left(\frac{x_1 + \cdots + x_n}{n}\right)^2.$$
$$E[X] = \frac{n+1}{2}; \quad Var(X) = \frac{n^2-1}{12}; \quad \textit{only when } x_i \in [1, 2, \ldots, n].$$

1.2 Bernoulli Random Variable

$$I = \begin{cases} 1 & \textit{if } X = 1, \\ 0 & \textit{if } X = 0. \end{cases}$$
$$E[I] = p; \quad Var(I) = p(1-p).$$

1.3 Binomial Random Variable

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \textit{for } k \in [0, n].$$
$$E[X] = np; \quad Var(X) = np(1-p).$$

1.4 Poisson Random Variable

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \textit{for } k \in \mathbb{Z}^{\geq}.$$
$$E[X] = \lambda; \quad Var(X) = \lambda.$$

1.5 Geometric Random Variable

$$P(X = n) = (1-p)^{n-1} p \quad \textit{for } n \in \mathbb{N}.$$
$$E[X] = \frac{1}{p}; \quad Var(X) = \frac{1-p}{p^2}; \quad P(X > n) = (1-p)^n; \quad P(X > n + k | X > k) = P(X > n).$$

2 Continuous Distributions

2.1 Uniform Random Variable

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \textit{if } x \in [\alpha, \beta], \\ 0 & \textit{otherwise.} \end{cases}$$

$$E[X] = \frac{\alpha+\beta}{2}; \quad Var(X) = \frac{(\beta-\alpha)^2}{12}.$$

2.2 Exponential Random Variable

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \textit{if } x \geq 0, \\ 0 & \textit{otherwise.} \end{cases}$$

$$E[X] = \frac{1}{\lambda}; \quad Var(X) = \frac{1}{\lambda^2}; \quad P(X \geq x) = e^{-\lambda x}; \quad P(X \geq x+y | X \geq y) = e^{-\lambda x}.$$

2.3 Normal Random Variable

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \textit{for } x \in \mathbb{R}.$$

$$E[X] = \mu; \quad Var(X) = \sigma^2; \quad Z \sim \textit{N(0,1)} = \frac{X-\mu}{\sigma}.$$

2.4 Gamma Random Variable

$$f(x) = \frac{\lambda}{\Gamma(\alpha)}(\lambda x)^{\alpha-1}e^{-\lambda x} \quad \textit{for } \alpha, \beta, x \in \mathbb{R}^+; \quad \Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt.$$

$$E[X] = \frac{\alpha}{\lambda}; \quad Var(X) = \frac{\alpha}{\lambda^2}.$$

3 Multivariate Distributions

3.1 Multivariate Normal Distribution

$$f(\vec{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^\top \Sigma^{-1}(\vec{x}-\vec{\mu})} \quad \textit{for } x \in \mathbb{R}^n.$$
$$E[X] = \vec{\mu}; \quad Var(X) = \Sigma.$$

Let random vector $\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ have a multivariate normal distribution and Z be the linear combination of the random vector with coefficients $\vec{A} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$. Then, $Z \sim N(\vec{A} \cdot \vec{\mu}, \vec{A}^\top \sum \vec{A})$. For bivariate random vector $\begin{bmatrix} X \\ Y \end{bmatrix}$ with $\vec{A} = \begin{bmatrix} a \\ b \end{bmatrix}$,

$$Var(Z) = a^2\sigma_X^2 + 2ab\rho\sigma_X\sigma_Y + b^2\sigma_Y^2.$$

3.2 Multinomial Distribution

$$P(\vec{x}) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} \quad \textit{for } \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix},\ x_i \in \mathbb{N},\ \sum_{i=1}^{k} x_i = n,\ p_i \in [0,1].$$
$$E[\vec{x}] = n\vec{p}; \quad Var(\vec{x}) = n\overrightarrow{p(1-p)}.$$

3.3 Random Sums

$$X = \xi_1 + \cdots + \xi_N; \quad \textit{where } N \textit{ is a R.V. and } \xi_i \textit{ are i.i.d. R.Vs.}$$
$$E[X] = E[\xi]E[N]; \quad Var(X) = E[N]Var(\xi) + E[\xi]^2 Var(N).$$

4 Theorems

4.1 Chebyshev's Inequality

$$P(|X - E[X]| \geq k) \leq \frac{Var(X)}{k^2}.$$

4.2 Markov Inequality

For a non-negative R.V. X,

$$P(X \geq k) \leq \frac{E[X]}{k}.$$

4.3 Central Limit Theorem

Let $S_n = X_1 + \cdots + X_n$. When n is large, S_n is approximately normally distributed with mean $n\mu$ and variance $n\sigma^2$.

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{\bar{S_n} - \mu}{\sigma/\sqrt{n}} \sim \mathit{N(0,1)} \quad \mathit{as}\ n \to \infty.$$

4.4 Weak Law of Large Numbers

Let $S_n = X_1 + \cdots + X_n$.

$$\lim_{n\to\infty} P(|\bar{S_n} - \mu| < \epsilon) = 1.$$

4.5 Strong Law of Large Numbers

Let $S_n = X_1 + \cdots + X_n$.

$$P(\lim_{n\to\infty} \bar{S_n} = \mu) = 1.$$

4.6 Moment Generating Function

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)dx.$$

5 Special Markov Chains

5.1 Two State Markov Chain

$$P = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \left\| \begin{array}{cc} 1-a & a \\ b & 1-b \end{array} \right\| \end{array}; \textit{ state 0, 1 are independent R.V. iff } a = 1-b.$$

$$P^n = \frac{1}{a+b}\begin{bmatrix} b & a \\ b & a \end{bmatrix} + \frac{(1-a-b)^n}{a+b}\begin{bmatrix} a & -a \\ -b & b \end{bmatrix}; \quad \lim_{n\to\infty} P^n = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \left\| \begin{array}{cc} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{array} \right\| \end{array}.$$

5.2 i.i.d. Markov Chain

$$\xi_1, \ldots, \xi_n \textit{ are i.i.d. R.Vs.}; \quad P(\xi = k) = a_k \textit{ for } k \in \mathbb{Z}^{\geq}.$$

5.2.1 Independent Random Variables

$$X_n = \xi; P = \left\| \begin{array}{cccc} a_0 & a_1 & a_2 & \cdots \\ a_0 & a_1 & a_2 & \cdots \\ a_0 & a_1 & a_2 & \cdots \\ a_0 & a_1 & a_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array} \right\|.$$

5.2.2 Successive Maxima

$$X_n = max(\xi_1, \ldots, \xi_n) \textit{ for } n \in \mathbb{N}.$$

$$P = \left\| \begin{array}{ccccc} A_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & A_1 & a_2 & a_3 & \cdots \\ 0 & 0 & A_2 & a_3 & \cdots \\ 0 & 0 & 0 & A_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array} \right\| \textit{where } A_k = P(X_{n+1} \leq x | X_n = x) = P(\xi_{n+1} \leq x).$$

Let T be the number of trials needed for $X_n = max(\xi_1, \ldots, \xi_n) \geq M$ for an arbitrary value M,

$$E[T] = \frac{1}{P(\xi \geq M)}.$$

Successive Maxima can be interpreted as a *Geometric(p)* random variable with $p = P(\xi \geq M)$.

5.2.3 Partial Sums

$$X_n = \xi_1 + \cdots + \xi_n \textit{ for } n \in \mathbb{N}.$$

$$P = \begin{Vmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ 0 & 0 & 0 & a_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{Vmatrix} \textit{ if } P(\xi = k) = a_k \textit{ for } k \in \mathbb{Z}^{\geq}.$$

$$P = \begin{Vmatrix} & \vdots & \vdots & \vdots & \vdots & \\ \cdots & a_{-1} & a_0 & a_1 & a_2 & \cdots \\ \cdots & a_{-2} & a_{-1} & a_0 & a_1 & \cdots \\ \cdots & a_{-3} & a_{-2} & a_{-1} & a_0 & \cdots \\ & \vdots & \vdots & \vdots & \vdots & \end{Vmatrix} \textit{ if } P(\xi = k) = a_k \textit{ for } k \in \mathbb{Z}.$$

5.2.4 Branching Processes

$$\begin{aligned} X_{n+1} &= \xi_1^{(n)} + \cdots + \xi_{X_n}^{(n)} \textit{ for } n \in \mathbb{N}; \quad X_0 = 1. \\ u_n &= \textit{probability of extinction by n-th generation} \\ &= \sum_{k=0}^{\infty} P(\xi = k)(u_{n-1})^k \textit{ for } n \in \mathbb{N}; \\ u_0 &= 0. \end{aligned}$$

From random sums, we know

$$E[X_{n+1}] = E[\xi]E[X_n]; \quad Var(X_{n+1}) = Var(\xi)E[X_n] + E[\xi]^2 Var(X_n).$$

After derivation,

$$E[X_n] = E[\xi]^n; \quad Var(X_n) = Var(\xi)E[\xi]^{n-1} \times \begin{cases} n & \textit{if } E[\xi] = 1, \\ \frac{1-E[\xi]^n}{1-E[\xi]} & \textit{otherwise.} \end{cases}$$

5.3 One-dimensional Random Walk

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{Vmatrix} r_0 & p_0 & 0 & 0 & \cdots \\ q_1 & r_1 & p_1 & 0 & \cdots \\ 0 & q_2 & r_2 & p_2 & \cdots \\ 0 & 0 & q_3 & r_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{Vmatrix} \end{array}.$$

5.3.1 Simple Random Walk

$$r_i = r = 0; \quad q_i = q = \frac{1}{2}; \quad p_i = p = \frac{1}{2}.$$

5.3.2 Gambler's Ruin

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ N-1 \\ N \end{array} \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & 3 & \cdots & N-1 & N \end{array} \\ \left\| \begin{array}{ccccccc} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ q & 0 & p & 0 & \cdots & 0 & 0 \\ 0 & q & 0 & p & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{array} \right\| \end{array}.$$

Let u_i be the probability of X_n reaching state 0 before state N given current state i.

$$u_i = \begin{cases} \frac{N-i}{N} & \textit{if } p = q = \frac{1}{2}, \\ \frac{(q/p)^i - (q/p)^N}{1-(q/p)^N} & \textit{if } p \neq q. \end{cases}$$

$$\lim_{N \to \infty} u_i = \begin{cases} 1 & \textit{if } p \leq q, \\ \left(\frac{q}{p}\right)^i & \textit{if } p > q. \end{cases}$$

Let T be the time needed to reach state 0.

$$E[T] = i(N-i) \ \textit{if } p = q = \frac{1}{2}.$$

5.4 Success Runs

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & \cdots \end{array} \\ \left\| \begin{array}{cccccc} p_0 & q_0 & 0 & 0 & 0 & \cdots \\ p_1 & r_1 & q_1 & 0 & 0 & \cdots \\ p_2 & 0 & r_2 & q_2 & 0 & \cdots \\ p_3 & 0 & 0 & r_3 & q_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array} \right\| \end{array}.$$

5.4.1 Two Outcomes Repeated Trials

$$r_i = r = 0; \quad q_i = q = \frac{1}{2}; \quad p_i = p = \frac{1}{2}.$$

5.4.2 Renewal Process

Model a light bulb's lifetime as a random variable ξ, where

$$P(\xi = k) = a_k \ \textit{for } k \in \mathbb{N},$$

and X_n as the age of the bulb in service at time n.

$$r_k = 0; \quad p_k = \frac{a_{k+1}}{a_{k+1} + a_{k+2} + \dots}; \quad q_k = 1 - p_k \ \textit{for } k \in \mathbb{Z}^{\geq}.$$

6 Estimators

6.1 Unbiasedness

$$Bias(\hat{\theta}) = E[\hat{\theta}] - \theta.$$

6.2 Efficiency

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of θ. $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2).$$

The relative efficiency of $\hat{\theta}_1$ to $\hat{\theta}_2$ is

$$\frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}.$$

6.3 Consistency

$\hat{\theta}$ is a consistent estimator of θ if

$$\lim_{n\to\infty} P(|\hat{\theta} - \theta| \geq \epsilon) = 0.$$

6.4 Sufficiency

An estimator $\hat{\theta}$ is sufficient for θ if the conditional distribution of $X_1, \ldots, X_n$ given $\hat{\theta}$ does not depend on θ; equivalently $\hat{\theta}$ is sufficient for θ if the likelihood function $L(\theta) = f(X; \theta)$ can be factorized as

$$L(\theta) = g(\hat{\theta}; \theta) b(X_1, \ldots, X_n).$$

6.5 Mean Squared Error (MSE)

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = Var(\hat{\theta}) + Bias(\hat{\theta})^2.$$

6.6 Minimum Variance Unbiased Estimator

6.6.1 Fishers Information

$$I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right].$$

6.6.2 Cramer-Rao Lower Bound

Let $\hat{\theta}$ be an unbiased estimator of θ, then

$$Var(\hat{\theta}) \geq \frac{1}{I(\theta)}.$$

If $Var(\hat{\theta}) = \frac{1}{I(\theta)}$, then $\hat{\theta}$ is a minimum variance unbiased estimator, aka asymptotically efficient.

6.7 Asymptotic Properties of Maximum Likelihood Estimators

6.7.1 Asymptotic Normality

$$\lim_{n\to\infty} \hat{\theta} \sim N(\theta, \frac{1}{nI(\theta)}).$$

6.7.2 Asymptotic Unbiasedness

$$\lim_{n\to\infty} E[\hat{\theta}] = \theta.$$

6.7.3 Asymptotic Efficiency

$$\lim_{n\to\infty} Var(\hat{\theta}) = \frac{1}{nI(\theta)}.$$