§1 Random Variables

Definition 1.1 (Memoryless).

$$\mathrm{P}(X \geq a+b \mid X \geq a) = \mathrm{P}(X \geq b)$$

Definition 1.2 (Variance).

$$\mathrm{Var}(X) := \mathrm{E}((X-\mu_X)^2)$$

Definition 1.3 (Covariance).

$$\mathrm{Cov}(X,Y) := \mathrm{E}((X-\mu_X)(Y-\mu_Y))$$

Definition 1.4 (Correlation).

$$\mathrm{Corr}(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

Definition 1.5 (Convergence in Probability). Given RVs $\{X_n\}$, we say that $X_n \xrightarrow{\mathrm{P}} Y$ when for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}(|X_n - Y| \geq \epsilon) = 0$$

1.1 Independence

X and Y being independent is equivalent to any of the following:

- $\mathrm{P}(X \in A, Y \in B) = \mathrm{P}(X \in A)\mathrm{P}(Y \in B)$
- Discrete RVs:
  - $p_{X,Y}(x,y) = p_X(x)p_Y(y)$
  - $p_{X|Y}(x \mid y) = p_X(x)$ if $p_Y(y) > 0$
- Abs Cts RVs:
  - $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
  - $f_{X|Y}(x \mid y) = f_X(x)$ if $f_Y(y) > 0$

1.2 Various Properties

Theorem 1.6 (Marginals).

- Discrete case: $p_X(x) = \sum_{y\in\mathbb{R}} p_{X,Y}(x,y)$
- Abs cts case: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$

Theorem 1.7. $\frac{d}{dx}F_X = f_X$, if f_X cts at x.

For general RV X, Y, constants a, b have:

- $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$
- $\mathrm{Var}(X+Y) = \mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y)$
- $\mathrm{Cov}(a,X) = 0$
- Covariance is symmetric and bilinear:
  $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
  $\mathrm{Cov}(X,aY+bZ) = a\mathrm{Cov}(X,Y)+b\mathrm{Cov}(X,Z)$
  Hence Cauchy-Schwarz applies, giving

$$|\mathrm{Corr}(X,Y)| \leq 1$$

For independent RV X, Y, have:

- $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$
- $\mathrm{Cov}(X,Y) = 0$ (converse implication false)

1.3 Functions of Random Variables

Let $h : \mathbb{R} \to \mathbb{R}$ be a function.

- Discrete: $\mathrm{E}(h(X)) = \sum_{x\in\mathbb{R}} h(x)p_X(x)$
  $h(X)$ must also be a discrete RV
- Abs Cts: $\mathrm{E}(h(X)) = \int_{-\infty}^{\infty} h(x)f_X(x)\,dx$

Theorem 1.8 (Abs Cts Change of Variables). Have X abs cts, and $h : \mathbb{R} \to \mathbb{R}$ differentiable and strictly increasing/decreasing. Then for $y = h(x)$, have

$$f_Y(y) = \frac{f_X(h^{-1}(y))}{|h'(h^{-1}(y))|}$$

thus Y is also an abs cts RV.

Theorem 1.9 (WLLN). Given sequence of RVs $\{X_i\}$, each with the same mean μ, variances bounded, then

$$\frac{\sum_{i=1}^n X_i}{n} = \frac{1}{n}S_n \xrightarrow{\mathrm{P}} \mu$$

Monte Carlo algorithms can be used to estimate the value of some definite integral I. Write the integral as an expectation of some RV, now generate values according to the RV's distribution and take the mean. WLLN says that $\frac{1}{n}S_n \xrightarrow{\mathrm{P}} I$.

1.4 Inequalities

Theorem 1.10 (Markov's Inequality). If $a > 0$, $X \geq 0$

$$\mathrm{P}(X \geq a) \leq \frac{\mathrm{E}(X)}{a}$$

Theorem 1.11 (Chebyshev's Inequality). If $a > 0$,

$$\mathrm{P}(|Y-\mu_Y| \geq a) \leq \frac{\mathrm{Var}(Y)}{a^2}$$

§2 Distributions

2.1 Discrete

In the below analogies, free throws are being repeatedly shot independently with probability θ of scoring.

- $\text{Bernoulli}(\theta)$: 'scores in 1 free throw'

$$p_X(k) = \begin{cases} \theta & k = 1 \\ 1-\theta & k = 0 \\ 0 & \text{otherwise} \end{cases}$$

- $\text{Binomial}(n, \theta)$: 'scores in n free throws'

$$p_X(k) = \begin{cases} \binom{n}{k}\theta^k(1-\theta)^{n-k} & k = \{0, \cdots, n\} \\ 0 & \text{otherwise} \end{cases}$$

- $\text{Geometric}(\theta)$: '# misses till score'

$$p_X(k) = \begin{cases} (1-\theta)^k\theta & k \in \{0, 1, \cdots\} \\ 0 & \text{otherwise} \end{cases}$$

- $\text{Poisson}(\lambda)$: approximation for $p_Y(k)$, where k is held constant, $Y \sim \text{Binom}(n, \frac{\lambda}{n})$ and n is large

$$p_X(k) = \begin{cases} e^{-\lambda}\frac{\lambda^k}{k!} & k \in \{0, 1, \cdots\} \\ 0 & \text{otherwise} \end{cases}$$

2.2 Absolutely Continuous

- $\text{Uniform}[a, b]$: 'perfect fairness'

$$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

- $\text{Exponential}(\lambda)$: can describe waiting time, 'continuous version of Geometric'

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- $\text{Normal}(\mu, \sigma^2)$: arises from CLT

$$f_Z(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

χ^2, t, and F distributions are related to the normal, not listed.

2.3 Normal Distribution

Density function of $\text{Normal}(\mu, \sigma^2)$ is given by $\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})$, where

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

Theorem 2.1. $W \sim \text{Normal}(\mu, \sigma^2)$ if and only if $W = \sigma Z + \mu$ where $Z \sim \text{Normal}(0, 1)$.

Theorem 2.2 (CLT)**.** Require $\{X_i\}$ to be iid, with the same finite mean μ and variance σ^2. Take $Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$. Have $\text{E}(Z_n) = 0$, $\text{Var}(Z_n) = 1$.

$$\lim_{n\to\infty} \text{P}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq z\right) = \Phi(z)$$

Can be written $F_{Z_n} \to \Phi$ (convergence in distribution).

2.4 Computations

Think of the $Y_i \sim \text{Binomial}(n, \theta)$ as the sum of $\sum_{i=1}^n X_i$, where $\{X_i\}$ is independent, and each $X_i \sim \text{Bernoulli}(\theta)$.

Distribution	Expectation	Variance
$\text{Bernoulli}(\theta)$	θ	$\theta(1-\theta)$
$\text{Binomial}(n, \theta)$	$n\theta$	$n\theta(1-\theta)$
$\text{Geometric}(\theta)$	$\frac{1-\theta}{\theta}$	$\frac{1-\theta}{\theta^2}$
$\text{Poisson}(\lambda)$	λ	λ
$\text{Uniform}[a, b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\text{Exponential}(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\text{Normal}(\mu, \sigma^2)$	μ	σ^2

To compute 95% confidence intervals, see that

$$\text{P}\left(\frac{1}{n}S_n - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \frac{1}{n}S_n + 1.96\frac{\sigma}{\sqrt{n}}\right) \approx 0.95$$

In the frequentist interpretation, S_n is the only random variable here. You can usually substitute σ with the sample standard deviation.