Random Variables Notes

Daniel Nichols

December 2018

1 Probability Space

Definition 1.1. (Sample Space). *The sample space* Ω *of an experiment is the set of all possible outcomes.*

Definition 1.2. (Event). *Given a sample space* Ω *we say that* A *is an event in* Ω *if* $A \subseteq \Omega$*.*

Definition 1.3. (Event Space). *An event space* $\mathcal{F}$ *over* Ω *is a set of events in* Ω *satisfying the following four properties,*

1. $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \implies (\Omega \setminus A) \in \mathcal{F}$
3. $A_i \in \mathcal{F} \implies \bigcup_i A_i \in \mathcal{F}$
4. $A_i \in \mathcal{F} \implies \bigcap_i A_i \in \mathcal{F}$

Since $\Omega \in \mathcal{F}$, then we must also have $\varnothing \in \mathcal{F}$. Typically we define $\mathcal{F} = 2^\Omega$.

Definition 1.4. (Probability Measure). *Given a sample space* Ω *and an event space* $\mathcal{F}$ *over* Ω*, a probability measure* $\mathbb{P}(\cdot)$ *is a function* $\mathbb{P} : \mathcal{F} \to \mathbb{R}^+$ *with the following two properties,*

1. $\mathbb{P}(\Omega) = 1$
2. $A_i \cap A_j = \varnothing$ (s.t. $i \neq j$) $\implies \mathbb{P}\left(\bigcup_i A_i\right) = \sum_i \mathbb{P}(A_i)$

Definition 1.5. (Disjoint Events). *Mutually exclusive events,*

$$i \neq j \implies A_i \cup A_j = \varnothing.$$

Definition 1.6. (Probability Space). *A probability space is a truple* $(\Omega, \mathcal{F}, \mathbb{P}(\cdot))$ *where* Ω *is a sample space,* $\mathcal{F}$ *is an event space over* Ω*, and* $\mathbb{P}(\cdot)$ *is a probability measure.*

2 Counting

Definition 2.1. (Permutations). *The number of permutations of an unordered list of n elements $\{a_0, a_1, \ldots, a_n\}$ is defined as P_n where*

$$P_n = n!$$

Definition 2.2. (Permute k from n with repetitions). *The number of ways to choose k elements from a set of n elements where repetitions are allowed and order matters is defined as $\overline{P}_k^n$ where*

$$\overline{P}_k^n = n^k.$$

Definition 2.3. (Permute k from n with no repetitions). *The number of ways to choose k elements from a set of n elements where repetitions are not allowed and order matters is defined as P_k^n where*

$$P_k^n = \frac{n!}{(n-k)!}.$$

Definition 2.4. (Choose k from n with no repetitions). *The number of ways to choose k elements from a set of n elements where repetitions are not allowed and order does not matter is defined as C_k^n where*

$$C_k^n = \binom{n}{k} = \frac{P_k^n}{k!} = \frac{n!}{k!(n-k)!}.$$

Definition 2.5. (Choose k from n with repetitions). *The number of ways to choose k elements from a set of n elements where repetitions are allowed and order does not matter is defined as $\overline{C}_k^n$ where*

$$\overline{C}_k^n = C_k^{k+n-1} = \binom{k+n-1}{k}.$$

3 Conditional Probability

Definition 3.1. (Conditional Probability). *For a probability space $(\Omega, \mathcal{F}, \mathbb{P}(\cdot))$ and two events A and B such that $\mathbb{P}(B) \neq 0$ we define the conditional probability of the event A given event B as*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Definition 3.2. (Independent Events). *In a probability space $(\Omega, \mathcal{F}, \mathbb{P}(\cdot))$the events $A_1, \ldots, A_n$ are jointly independent if*

$$\mathbb{P}\left(\bigcap_i A_i\right) = \prod_i \mathbb{P}(A_i).$$

Definition 3.3. (Conditional Probability of Independent Events). *If events A and B are independent, then*

$$\mathbb{P}(A|B) = \mathbb{P}(A)$$

$$\mathbb{P}(B|A) = \mathbb{P}(B).$$

Definition 3.4. (Law of Total Probability). *Let A and $B_1, \ldots, B_n$ be events in the probability space $(\Omega, \mathcal{F}, \mathbb{P}(\cdot))$. Assume that $\mathbb{P}(B_i) > 0$ $(\forall i \in \mathbb{N})$, $B_i \cap B_j$ $(\forall i \neq j)$, and $\bigcup_i B_i = \Omega$. This implies*

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Definition 3.5. (Bayes' Theorem). *Let A and B be two events in a probability space $(\Omega, \mathcal{F}, \mathbb{P}(\cdot))$. Then it follows that*

$$\mathbb{P}(A|B) = \mathbb{P}(B|A)\frac{\mathbb{P}(A)}{\mathbb{P}(B)}.$$

4 Random Variables

Definition 4.1. (Random Variable). *In a probability space $(\Omega, \mathcal{F}, \mathbb{P}(\cdot))$, a random variable X is a function from the sample space Ω to a domain $\mathbb{D}$, i.e. $X : \Omega \to \mathbb{D}$. If $\mathbb{D} = \mathbb{R}$, then we say X is a continuous random variable. Otherwise, if $\mathbb{D} \subset \mathbb{R}$, then we X is a discrete random variable.*

Definition 4.2. (Probability Cumulative Function). *Given a random variable X we define its Probability Cumulative Function $F_X : \mathbb{R} \to \mathbb{D}$ as*

$$F_X(x) := \mathbb{P}(X \leq x).$$

The PCF of a random variable is a monotonic increasing function.

Definition 4.3. (PCF Over an Interval). *For a random variable X and an interval $(a, b]$, we have*

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).$$

Definition 4.4. (Probability Mass Function). *The Probability Mass Function of a discrete random variable X with PCF $F_X(x)$ is given by $\mathbb{P}(X = x)$ such that*

$$\sum_{x \in \mathbb{D}} \mathbb{P}(X = x) = 1.$$

Definition 4.5. (Probability Density Function). *The Probability Density Function of a random variable X with PCF $F_X(x)$ is given by the function $f_x : \mathbb{R} \to \mathbb{R}^+$ such that*

$$f_x(x) = \frac{d}{dx}F_X(x).$$

The PDF has two key properties, (1) non-negativity and (2) normalization.

1. $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$
2. $\int_{\mathbb{R}} f_X(x)\, dx = 1$

We can also use the properties of PMF's and the fundamental theorem of calculus to write,

$$\mathbb{P}(a < X \leq b) = \int_a^b f_X(x)\, dx.$$

5 Expectation

Definition 5.1. (Expectation of a Discrete Random Variable). *Given a discrete random variable X with domain $\mathbb{D}$ and a function $\varphi : \mathbb{D} \to \mathbb{D}$, we define the expectation of $\varphi(X)$ as*

$$\mathbb{E}[\varphi(X)] := \sum_{x \in \mathbb{D}} \varphi(x)\mathbb{P}(X = x).$$

Definition 5.2. (Expectation of a Continuous Random Variable). *Given a continuous random variable X with PDF f_X and a function $\varphi : \mathbb{R} \to \mathbb{R}$ we define the expectation of $\varphi(X)$ as*

$$\mathbb{E}[\varphi(X)] := \int_{\mathbb{R}} \varphi(x) f_X(x)\, dx.$$

Expectation has some nice properties we can use,

$$\mathbb{E}[\varphi_1(X_1) + \varphi_2(X_2)] = \mathbb{E}[\varphi_1(X_1)] + \mathbb{E}[\varphi_2(X_2)]$$
$$\mathbb{E}[\alpha\varphi(X)] = \alpha\mathbb{E}[\varphi(X)] \quad (\forall \alpha \in \mathbb{R}).$$

Definition 5.3. (Moments). *Given a random variable X we define its k-th moment as $\mathbb{E}[X^k]$, for $k \in \mathbb{N}$.*

Definition 5.4. (Mean). *Given a random variable X its mean, μ_X, is defined as its first moment, namely*

$$\mu_X := \mathbb{E}[X].$$

Definition 5.5. (Central Moment). *Given a random variable X and $k \in \mathbb{N}$ we define its k-th central moment as*

$$\mathbb{E}[(X - \mu_X)^k].$$

Definition 5.6. (Variance). *Given a random variable X its variance, σ_X^2, is defined as the 2nd central moment. Namely,*

$$\sigma_X^2 = \mathbb{E}[(X - \mu_X)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Definition 5.7. (Markov Inequality). *Given a random variable* X *such that* $\mathbb{P}(X < 0) = 0$*, then we have*

$$\mathbb{P}(X \geq k) \leq \frac{\mathbb{E}[X]}{k} \quad (\forall k > 0).$$

Definition 5.8. (Chebyshev Inequality). *Given a random variable* X *with* $\mathbb{E}[X] = \mu_X$ *and variance* σ_X^2*, then we have*

$$\mathbb{P}(|X - \mu_X| \geq k\sigma_X) \leq \frac{1}{k^2} \quad (\forall k > 0).$$

6 Distributions

Definition 6.1. (Uniform Distribution). *if* $X \sim \mathcal{U}(a, b)$ *then* X *takes values uniformly on* $[a, b]$ *and*

$$\mathbb{P}(X = x) = \begin{cases} \frac{1}{b-a} & if\ a \leq x < b \\ 0 & if\ x < a\ or\ b \leq x \end{cases}$$

$$\mu_X = \frac{a + b}{2}$$
$$\sigma_X^2 = \frac{(b - a)^2}{12}.$$

Definition 6.2. (Bernoulli Distribution). *If* $X \sim Bernoulli(p)$*, then* p *is called the success parameter and* X *can only take values of* 0 *or* 1*. The probabilities are given by*

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

$$\mu_X = p$$
$$\sigma_X^2 = p(1 - p).$$

Definition 6.3. (Binomial Distribution). *If* $X \sim Binomial(n, p)$*, where* p *and* n *are the probability of success and the number of trials, then*

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$
$$\mu_X = np$$
$$\sigma_X^2 = np(1 - p).$$

Definition 6.4. (Geometric Distribution). *If $X \sim Geom(p)$, where p is the probability of success, then*

$$\mathbb{P}(X = k) = (1-p)^k p$$
$$\mu_X = \frac{p}{1-p}$$
$$\sigma_X^2 = \frac{p}{(1-p)^2}.$$

Definition 6.5. (Exponential Distribution). *If $X \sim Exp(\lambda)$ then we say X has a decay rate of λ and*

$$\mathbb{P}(X = x^+) = \lambda e^{-\lambda x}$$
$$\mu_X = \frac{1}{\lambda}$$
$$\sigma_X^2 = \frac{1}{\lambda^2}.$$

Definition 6.6. (Poisson Distribution). *If $X \sim Poisson(\Lambda)$ then X is used to describe the number of events occurring independently in a fixed amount of time with intervals between them distributed exponentially. It is given by*

$$\mathbb{P}(X = k) = \frac{\Lambda^k}{k!} e^{-\Lambda}$$
$$\mu_X = \Lambda$$
$$\sigma_X^k = \Lambda.$$

Definition 6.7. (Hypergeometric Distribution). *If $X \sim Hypergeom(N, K, n)$, then X is hypergeometrically distributed. The hypergeometric distribution is useful for a population size N with K successes in it and n removed without replacement and we want to know how likely k out of the n were successful. It is given by*

$$\mathbb{P}(X = k) = \frac{\binom{N}{K}\binom{N-K}{n-k}}{\binom{N}{n}}$$
$$\mu_X = n\frac{K}{N}$$
$$\sigma_X^2 = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}.$$

Definition 6.8. (Gaussian Distribution). *If $X \sim \mathcal{N}(\mu, \sigma^2)$ then it is given by*

$$\mathbb{P}(X = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$
$$\mu_X = \mu$$
$$\sigma_X^2 = \sigma^2.$$

7 Predictive Statistics

Definition 7.1. (Biased Estimate). *μ_X is an unbiased estimate for $\mathbb{E}[X]$ if $\mathbb{E}[\mu_X - g(X_0, \ldots, X_n)] = 0$, where $g : \mathbb{D}^n \to \mathbb{R}$. It is biased otherwise.*

Estimation of X and Y if $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. We can use the formula

$$\hat{X} = \alpha Y.$$

To find an optimal value for α we need to compute

$$\min_{\alpha} \mathbb{E}[(X - \alpha Y)^2].$$

This gives us

$$\alpha = \frac{R_{XY}}{\sigma_Y^2}.$$

where $R_{XY} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$.

If $\mathbb{E}[X] \neq 0$ and $\mathbb{E}[Y] \neq 0$, then we need an additional bias in our estimate,

$$\hat{X} = \alpha Y + \beta.$$

Now we need to find α and β by computing

$$\min_{\alpha,\beta} \mathbb{E}[(X - (\alpha Y + \beta))^2].$$

This gives us values of

$$\alpha = \frac{R_{XY}}{\sigma_Y^2}, \quad \beta = \mathbb{E}[X] - \frac{R_{XY}}{\sigma_Y^2}\mathbb{E}[Y].$$

The correlation coefficient for X and Y is given by

$$\rho_{XY} = \frac{R_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}}.$$

If these values are unknown, then we can use $\mathbb{E}[X] \approx \hat{\mu}_X$, $\mathbb{E}[Y] \approx \hat{\mu}_Y$, $\sigma_Y^2 \approx \hat{\sigma}_Y^2$, and $R_{XY} \approx \hat{R}_{XY}$. We can define these as,

$$\begin{aligned}
\hat{\mu}_X &= \frac{1}{n} \sum_i X_i \\
\hat{\mu}_Y &= \frac{1}{n} \sum_i Y_i \\
\hat{\sigma}_Y^2 &= \frac{1}{n-1} \sum_i (Y_i - \hat{\mu}_Y)^2 \\
\hat{R}_{XY} &= \frac{1}{n-1} \sum_i (X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y).
\end{aligned}$$