Basics of Probability

Eklavya Sharma

Definition 1 (Probability Space). *A probability space is a triple $(\Omega, \mathcal{F}, P)$ where*

- *Ω is the sample space, also called the set of all outcomes.*
- *$\mathcal{F}$ is a σ-algebra over Ω. $\mathcal{F}$ is called the set of all events.*
- *$P : \mathcal{F} \mapsto [0, 1]$ is a measure over $(\Omega, \mathcal{F})$ (i.e., P is σ-additive) such that $P(\Omega) = 1$. P is called the probability measure.*

Theorem 1 (Inclusion-Exclusion Principle).

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \cdots < i_k \le n} \Pr(A_{i_1} \cap \cdots \cap A_{i_k}).$$

Theorem 2. *For randvars X and Y, $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$.*

Theorem 3. *For independent randvars $X_1, \ldots, X_n$, $\mathrm{E}(X_1 \ldots X_n) = \mathrm{E}(X_1) \ldots \mathrm{E}(X_n)$.*

Theorem 4. *For a non-negative randvar X,*

$$\mathrm{E}(X) = \begin{cases} \displaystyle\sum_{i=0}^{\infty} \Pr(X > i) & \text{if } X \text{ is discrete} \\ \displaystyle\int_0^{\infty} \Pr(X > x) dx & \text{if } X \text{ is continuous} \end{cases}.$$

Definition 2.

$$\begin{aligned} \mathrm{Cov}(X, Y) &:= \mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) \\ \mathrm{Var}(X) &:= \mathrm{Cov}(X, X) = \mathrm{E}((X - \mathrm{E}(X))^2) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 \end{aligned}$$

Theorem 5.

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le n} \mathrm{Cov}(X_i, X_j).$$

Theorem 6. *Let $\mathrm{MGF}_t(X) := \mathrm{E}(e^{tX})$. Then MGF_t uniquely determines X's CDF.*

Theorem 7 (Change of variables). *Let $X \in \mathbb{R}^n$ be a continuous random vector. Let $g : \mathbb{R}^n \mapsto \mathbb{R}^n$ be a bijective function having continuous partial derivatives. Then $f_{g(X)}(y) = f_X(x)|J_g(x)|^{-1}$, where $x := g^{-1}(y)$ and J_g is the Jacobian of g (i.e., $J_g(x)[i, j] := \partial g(x)_i / \partial x_j$).*

Definition 3. *Let $A = [A_1, A_2, \ldots]$ be an infinite sequence of events. Then*

$$\mathrm{io}(A) = \lim_{m\to\infty} \bigcup_{i=m}^{\infty} A_i = \bigcap_{m=1}^{\infty} \bigcup_{i=m}^{\infty} A_i \qquad \mathrm{ae}(A) = \lim_{m\to\infty} \bigcap_{i=m}^{\infty} A_i = \bigcup_{m=1}^{\infty} \bigcap_{i=m}^{\infty} A_i.$$

$\mathrm{io}(A)$ *are the outcomes in* Ω *for which infinitely many events in* A *happen.* $\mathrm{ae}(A)$ *are the outcomes in* Ω *for which all except finitely many events in* A *happen.*

Lemma 8 (Borel-Cantelli)**.** $\sum_{i=1}^{\infty} \Pr(A_i) \neq \infty \implies \Pr(\mathrm{io}(A)) = 0$.

Lemma 9. *(Events in A are independent and $\sum_{i=1}^{\infty} \Pr(A_i) = \infty$)* $\implies \Pr(\mathrm{io}(A)) = 1$.

1 Probability Distributions

Table 1: Discrete Probability Distributions

Distribution	$\Pr(X = x)$	$\mathrm{E}(X)$	$\mathrm{Var}(X)$	$\mathrm{MGF}_t(X)$
Bernouilli(p)	$p^x(1-p)^{1-x}$	p	$p(1-p)$	$pe^t + 1 - p$
Binomial(n, p)	$\binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$	$(pe^t + 1 - p)^n$
Geometric(p)	$(1-p)^{x-1}p$	$1/p$	$(1-p)/p^2$	$\dfrac{pe^t}{1-(1-p)e^t}$
Poisson(λ)	$e^{-\lambda}\lambda^x/x!$	λ	λ	$\exp(\lambda(e^t - 1))$

Theorem 10 (Poisson approximates Binomial)**.** *Let $\lambda \in \mathbb{R}_{\geq 0}$ and $k \in \mathbb{Z}_{\geq 0}$ be constants. Let $X_n \sim \mathrm{Binom}(n, \lambda/n)$. Then $\lim_{n\to\infty} \Pr(X_n = k) = e^{-\lambda}\lambda^k/k!$.*

Table 2: Continuous Probability Distributions

Distribution	$f_X(x)$	$\mathrm{E}(X)$	$\mathrm{Var}(X)$	$\mathrm{MGF}_t(X)$
Uniform(a, b)	$\dfrac{\mathbf{1}(a \leq x \leq b)}{b-a}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{e^{bt} - e^{at}}{(b-a)t}$
Exponential(λ)	$\lambda e^{-\lambda x}$	$1/\lambda$	$1/\lambda^2$	$\lambda/(\lambda - t)$
Gamma(n, λ)	$\dfrac{(\lambda x)^{n-1}}{(n-1)!}\lambda e^{-\lambda x}$	n/λ	n/λ^2	$\left(1 - \dfrac{t}{\lambda}\right)^{-n}$
Normal(μ, σ^2)	$\dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2	$\exp(\mu t + \sigma^2 t^2/2)$

Theorem 11 (Scaling normal)**.** $X \sim N(\mu, \sigma^2) \implies aX + b \sim N(a\mu + b, a^2\sigma^2)$.

1.1 Sum of Random Variables

Theorem 12 (Convolution).

$$f_{X+Y}(z) = \begin{cases} \sum_{y \in D} f_{X,Y}(z-y, y) = \sum_{x \in D} f_{X,Y}(x, z-x) & \textit{discrete} \\ \int_{-\infty}^{\infty} f_{X,Y}(z-y, y)dy = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)dx & \textit{continuous} \end{cases}.$$

Theorem 13. *Let* $X_1, \ldots, X_n$ *be independent. Then* $\operatorname{MGF}_t\left(\sum_{i=1}^n X_i\right) = \prod_{i=1}^n \operatorname{MGF}_t(X_i)$.

Theorem 14. *Let* $X_1, \ldots, X_n$ *be independent. Let* $Y := \sum_{i=1}^n X_i$. *Then*

- $X_i \sim \operatorname{Bernouilli}(p) \implies Y \sim \operatorname{Binomial}(n, p)$.
- $X_i \sim \operatorname{Poisson}(\lambda_i) \implies Y \sim \operatorname{Poisson}\left(\sum_{i=1}^n \lambda_i\right)$.
- $X_i \sim \operatorname{Exponential}(\lambda) \implies Y \sim \operatorname{Gamma}(n, \lambda)$.

2 Inequalities and Limits

Theorem 15 (Markov). *For non-negative randvar* X, $\Pr(X \geq a) \leq \operatorname{E}(X)/a$.

Theorem 16 (Chebyshev). $\Pr(|X - \operatorname{E}(X)| \geq a) \leq \dfrac{\operatorname{Var}(X)}{a^2}$.

Theorem 17 (One-sided Chebyshev).

$$\Pr(X - \operatorname{E}(X) \geq a) \leq \frac{\operatorname{Var}(X)}{\operatorname{Var}(X) + a^2} \qquad \Pr(X - \operatorname{E}(X) \leq -a) \leq \frac{\operatorname{Var}(X)}{\operatorname{Var}(X) + a^2}$$

Theorem 18 (Strong law of large lumbers). *Let* $X_1, X_2, \ldots$ *be IID randvars having mean* μ. *Let* $Y_n := \frac{1}{n}\sum_{i=1}^n X_i$. *Let*

$$E := \left\{\omega \in \Omega : \lim_{n\to\infty} Y_n(\omega) = \mu\right\}.$$

Then $\Pr(E) = 1$.

Definition 4. *Let* Z *be a random variable and* $S := [X_1, X_2, \ldots]$ *be an infinite sequence of random variables. We say that* S *converges to* Z *if* $\lim_{n\to\infty} F_{X_n}(x) = F_Z(x)$ *for all* $x \in \mathbb{R}$ *where* F_Z *is continuous.*

Theorem 19 (Central Limit Theorem). *Let* $X_1, X_2, \ldots$ *be IID randvars having mean* μ *and variance* σ^2. *Let* $Y_n := \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mu\right)$. *Then* $[Y_1, Y_2, \ldots]$ *converges to* $N(0, \sigma^2)$.

Theorem 20 (Jensen's inequality). *If* X *is a random variable and* f *is a convex function, then* $f(\operatorname{E}(X)) \leq \operatorname{E}(f(X))$.

Theorem 21 (Cauchy-Schwarz inequality). *For random variables* X *and* Y, $|\operatorname{E}(XY)|^2 \leq \operatorname{E}(X^2)\operatorname{E}(Y^2)$ *and* $|\operatorname{Cov}(X, Y)|^2 \leq \operatorname{Var}(X)\operatorname{Var}(Y)$.

3 Conditional Probability

Definition 5. *Let X and Y be continuous randvars. Let $f_{Y|X}(y \mid x) := f_{X,Y}(x,y)/f_X(x)$. $f_{Y|X}$ is called the* density function of Y conditioned on X. *Then* $\mathrm{E}(Y \mid X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x) dx$.

Definition 6. *Let X and Y be randvars and A be an event. Let $g(x) := \Pr(A \mid X = x)$ and $h(x) := \mathrm{E}(Y \mid X = x)$. Then $\Pr(A \mid X) := g(X)$ and $\mathrm{E}(Y \mid X) := h(X)$.*

Theorem 22. $\mathrm{E}(\Pr(A \mid X)) = \Pr(A)$ *and* $\mathrm{E}(\mathrm{E}(Y \mid X)) = \mathrm{E}(Y)$.

Theorem 23. $\mathrm{Var}(Y) = \mathrm{E}(\mathrm{Var}(Y \mid X)) + \mathrm{Var}(\mathrm{E}(Y \mid X))$.

4 Binomial Coefficient

The binomial coefficient $\binom{n}{k}$ is the number of subsets of $\{1, 2, \ldots, n\}$ of size k, where $n \in \mathbb{Z}_{\geq 0}$ and $k \in \mathbb{Z}$.

- $\displaystyle \binom{n}{k} = \binom{n}{n-k} = \begin{cases} 0 & \text{if } k < 0 \text{ or } k > n \\ \frac{n!}{k!(n-k)!} & \text{if } 0 \leq k \leq n \end{cases}.$
- Additive recursion: For $n \geq 1$, $\displaystyle \binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n+1}{k+1} - \binom{n}{k+1}$.
- Decrement: For $n \geq 1$, $\displaystyle \binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1} = \frac{n}{n-k}\binom{n-1}{k} = \frac{n-k+1}{k}\binom{n}{k-1}$.
- Sum 1: $\displaystyle \sum_{i=k}^{n} \binom{i}{k}\binom{n}{i} x^i = \binom{n}{k} x^k (1+x)^{n-k}$. Set $k = 0$ to get $\displaystyle \sum_{i=0}^{n} \binom{n}{i} x^i = (1+x)^n$.
- Sum 2: $\displaystyle \sum_{i=0}^{p} \binom{m}{i}\binom{n}{p-i} = \binom{m+n}{p}$.
- Sum 3: $\displaystyle \sum_{i=k}^{n-b} \binom{i}{k}\binom{n-i}{b} = \binom{n+1}{k+b+1}$. Set $b = 0$ to get $\displaystyle \sum_{i=k}^{n} \binom{i}{k} = \binom{n+1}{k+1}$.

5 Other useful results

$$\forall x \in \mathbb{R}, \quad e^x \geq 1 + x.$$

$$\forall x > 0, \quad \frac{x-1}{x} \leq \ln x \leq x - 1.$$

$$\forall n \geq 1, \quad \left(\sum_{i=1}^{n} \frac{1}{i}\right) - \ln n \in [1/n, 1].$$

Stirling's approximation: For $n \geq 1$, $\displaystyle \frac{n!}{n^{n+\frac{1}{2}} e^{-n}} \in [\sqrt{2\pi}, e]$.

$$\forall a \in \mathbb{Z}, \forall b \in \mathbb{Z}_{>0}, \quad \left\lceil \frac{a}{b} \right\rceil = \left\lfloor \frac{a-1}{b} \right\rfloor + 1 \quad \text{and} \quad \left\lfloor \frac{a}{b} \right\rfloor = \left\lceil \frac{a+1}{b} \right\rceil - 1.$$

Theorem 24 (Generalization of Geometric series)**.** *For $0 \leq a \leq b$,*

$$\sum_{i=0}^{\infty} \binom{b+i}{a} p^i = \frac{1}{1-p} \sum_{i=0}^{a} \binom{b}{i} \left(\frac{p}{1-p}\right)^{a-i}.$$

On setting $b = a$, we get

$$\sum_{i=0}^{\infty} \binom{a+i}{a} p^i = \frac{1}{(1-p)^{a+1}}.$$