

Basics of Probability

Eklavya Sharma

Definition 1 (Probability Space). A probability space is a triple (Ω, \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 (Ω, \mathcal{F}) (i.e., P is σ -additive) such that $P(\Omega) = 1$. P is called the probability measure.

Theorem 1 (Inclusion-Exclusion Principle).

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

Theorem 2. For randvars X and Y , $E(X + Y) = E(X) + E(Y)$.

Theorem 3. For independent randvars X_1, \dots, X_n , $E(X_1 \dots X_n) = E(X_1) \dots E(X_n)$.

Theorem 4. For a non-negative randvar X ,

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

Definition 2.

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

Theorem 5.

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

Theorem 6. Let $\text{MGF}_t(X) := E(e^{tX})$. 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 $A = [A_1, A_2, \dots]$ be an infinite sequence of events. Then

$$\text{io}(A) = \lim_{m \rightarrow \infty} \bigcup_{i=m}^{\infty} A_i = \bigcap_{m=1}^{\infty} \bigcup_{i=m}^{\infty} A_i \quad \text{ae}(A) = \lim_{m \rightarrow \infty} \bigcap_{i=m}^{\infty} A_i = \bigcup_{m=1}^{\infty} \bigcap_{i=m}^{\infty} A_i.$$

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

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

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

1 Probability Distributions

Table 1: Discrete Probability Distributions

Distribution	$\Pr(X = x)$	$E(X)$	$\text{Var}(X)$	$\text{MGF}_t(X)$
Bernoulli(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)	$(1-p)^{x-1}p$	$1/p$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1 - (1-p)e^t}$
NegBinom(n, p)	$\binom{x-1}{n-1} p^n(1-p)^{x-n}$	n/p	$\frac{n(1-p)}{p^2}$	$\left(\frac{pe^t}{1 - (1-p)e^t}\right)^n$
Poisson(λ)	$e^{-\lambda} \lambda^x / x!$	λ	λ	$\exp(\lambda(e^t - 1))$

Table 2: Continuous Probability Distributions

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

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

Theorem 11 (Binomial over Poisson). Let $N \sim \text{Poisson}(\lambda)$ and $M \mid N \sim \text{Binom}(N, p)$. Then $M \sim \text{Poisson}(\lambda p)$.

Theorem 12 (Poisson decomposition). Let X_1, X_2, \dots be IID randvars, where $\Pr(X_j = i) = p_i$ for $i \in [k]$ and all j , and $\sum_{i=1}^k p_i = 1$. Let $N \sim \text{Poisson}(\lambda)$ where $\{N, X_1, X_2, \dots\}$ is independent. Let $N_i := \sum_{j=1}^N \mathbf{1}(X_j = i)$. Then $\{N_1, \dots, N_k\}$ is independent and $N_i | N \sim \text{Binom}(N, p_i)$.

Theorem 13 (Scaling normal). $X \sim N(\mu, \sigma^2) \implies aX + b \sim N(a\mu + b, a^2\sigma^2)$.

Theorem 14. Let $X \sim \text{Expo}(\lambda)$. Then $E(X^k e^{-\mu X}) = \lambda k! / (\lambda + \mu)^{k+1}$.

Proof sketch. Use $E(f(X)) = \int_0^\infty f(x) \lambda e^{-\lambda x} dx$, then use integration by parts k times. \square

Theorem 15 (Competing exponentials). Let X_1, \dots, X_n be independent ranvars, where $X_i \sim \text{Expo}(\lambda_i)$. Let $Z := \min_{i=1}^n X_i$ and E be the event $X_1 < X_2 < \dots < X_n$. Let $\beta := \lambda_1 + \dots + \lambda_n$. Then

- $\Pr(X_i = Z) = \lambda_i / \beta$.
- $Z \sim \text{Expo}(\beta)$.
- E and Z are independent.

1.1 Sum of Random Variables

Theorem 16 (Convolution).

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

Theorem 17. Let X_1, \dots, X_n be independent. Then $\text{MGF}_t(\sum_{i=1}^n X_i) = \prod_{i=1}^n \text{MGF}_t(X_i)$.

Theorem 18. Let X_1, \dots, X_n be independent. Let $Y := \sum_{i=1}^n X_i$. Then

- $X_i \sim \text{Bernoulli}(p) \implies Y \sim \text{Binomial}(n, p)$.
- $X_i \sim \text{Poisson}(\lambda_i) \implies Y \sim \text{Poisson}(\sum_{i=1}^n \lambda_i)$.
- $X_i \sim \text{Exponential}(\lambda) \implies Y \sim \text{Gamma}(n, \lambda)$.
- $X_i \sim \text{Geometric}(p) \implies Y \sim \text{NegBinom}(n, p)$.

2 Inequalities and Limits

Theorem 19 (Markov). For non-negative randvar X , $\Pr(X \geq a) \leq E(X)/a$.

Theorem 20 (Chebyshev). $\Pr(|X - E(X)| \geq a) \leq \frac{\text{Var}(X)}{a^2}$.

Theorem 21 (One-sided Chebyshev).

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

Theorem 22 (Strong law of large numbers). Let X_1, X_2, \dots be IID randvars having mean μ . Let $Y_n := \frac{1}{n} \sum_{i=1}^n X_i$. Let

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

Then $\Pr(E) = 1$.

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

Theorem 23 (Central Limit Theorem). Let X_1, X_2, \dots 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 $[Y_1, Y_2, \dots]$ converges to $N(0, \sigma^2)$.

Theorem 24 (Jensen's inequality). If X is a random variable and f is a convex function, then $f(E(X)) \leq E(f(X))$.

Theorem 25 (Cauchy-Schwarz inequality). For random variables X and Y , $|E(XY)|^2 \leq E(X^2)E(Y^2)$ and $|\text{Cov}(X, Y)|^2 \leq \text{Var}(X)\text{Var}(Y)$.

3 Conditional Probability

Theorem 26. Let X and Y be randvars (either of them can be discrete or continuous). Let f_X and f_Y be their distribution functions (either PMF or PDF), respectively. Let $f_{X,Y}$ be their joint distribution function. Let g_x be the distribution function of Y conditioned on $X = x$. Then $g_x(y) = f_{X,Y}(x, y)/f_X(x)$. We denote $g_x(y)$ by $f_{Y|X}(y | x)$.

Definition 5. Let X and Y be randvars and A be an event. Let $g(x) := \Pr(A | X = x)$ and $h(x) := E(Y | X = x)$. Then $\Pr(A | X) := g(X)$ and $E(Y | X) := h(X)$.

Theorem 27. $E(\Pr(A | X)) = \Pr(A)$ and $E(E(Y | X)) = E(Y)$.

Theorem 28. $\text{Var}(Y) = E(\text{Var}(Y | X)) + \text{Var}(E(Y | X))$.

4 Binomial Coefficient

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

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

- Sum 2: $\sum_{i=0}^p \binom{m}{i} \binom{n}{p-i} = \binom{m+n}{p}$.
- Sum 3: $\sum_{i=k}^{n-b} \binom{i}{k} \binom{n-i}{b} = \binom{n+1}{k+b+1}$. Set $b = 0$ to get $\sum_{i=k}^n \binom{i}{k} = \binom{n+1}{k+1}$.

5 Other useful results

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

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

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

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

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

Theorem 29 (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

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