Probability Review Notes

1 Axioms of Probability (Kolmogorov)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space where Ω is the sample space, $\mathcal{F}$ a σ-algebra, and $\mathbb{P} : \mathcal{F} \to [0,1]$ a probability measure. For all events $A, B \in \mathcal{F}$:

(a) **Nonnegativity:** $\mathbb{P}(A) \geq 0$.

(b) **Normalization:** $\mathbb{P}(\Omega) = 1$.

(c) **σ-additivity:** If $A_1, A_2, \ldots$ are pairwise disjoint, then $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Consequences: $\mathbb{P}(\varnothing) = 0$; $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$; if $A \subseteq B$ then $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$; inclusion–exclusion: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

2 Conditional Probability and Bayes' Theorem

Definition 1 (Conditional Probability)**.** *For* $\mathbb{P}(B) > 0$*,*

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

Law of Total Probability: If $\{B_i\}_{i \in I}$ is a partition with $\mathbb{P}(B_i) > 0$,

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

Bayes' Theorem:

$$\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}{\sum_i \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}.$$

Odds form: Prior odds $\times$ likelihood ratio $=$ posterior odds.

3 Independence

Definition 2 (Events)**.** *Events A, B are* independent *if* $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$*. A family* $\{A_i\}_{i \in I}$ *is* mutually (jointly) independent *if for every finite distinct* $i_1, \ldots, i_k$*,*

$$\mathbb{P}\left(\bigcap_{\ell=1}^{k} A_{i_\ell}\right) = \prod_{\ell=1}^{k} \mathbb{P}(A_{i_\ell}).$$

They are pairwise independent *if* $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ *for all* $i \neq j$*.*

Important: Joint (mutual) independence $\Rightarrow$ pairwise independence, but not conversely. $\Rightarrow$

4 r.v.

Definition 3 (Random Variables)**.** *A random variable (r.v.)* $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ *is a measurable function:* $\{\omega : X(\omega) \le x\} \in \mathcal{F}$ *for all* $x \in \mathbb{R}$*.*

Distribution of X: $F_X(x) = \mathbb{P}(X \le x)$; for discrete X, pmf $p_X(x) = \mathbb{P}(X = x)$; for continuous X, pdf $f_X = \frac{d}{dx} F_X$ (where it exists).

5 Important Discrete Random Variables

Below $k \in \{0, 1, 2, \dots\}$ unless stated.

Bernoulli(p)

$\mathbb{P}(X = 1) = p,\ \mathbb{P}(X = 0) = 1 - p.$

$$\mathbb{E}[X] = p, \quad \mathrm{Var}(X) = p(1-p), \quad M_X(t) = 1 - p + pe^t.$$

Binomial(n, p)

Sum of n i.i.d. Bernoulli(p):

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \mathbb{E}[X] = np, \quad \mathrm{Var}(X) = np(1-p).$$

Geometric(p) (number of trials to first success)

Support $\{1, 2, \dots\}$:

$$\mathbb{P}(X = k) = (1-p)^{k-1} p, \quad \mathbb{E}[X] = \frac{1}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2}.$$

Memoryless: $\mathbb{P}(X > m + n \mid X > m) = (1-p)^n$.

Poisson(λ)

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad \mathbb{E}[X] = \mathrm{Var}(X) = \lambda.$$

Poisson thinning/superposition; Poisson limit of Binomial: if $n \to \infty$, $p \to 0$, $np \to \lambda$.

Discrete Uniform on $\{a, \dots, b\}$

$p_X(k) = 1/(b - a + 1)$.

$$\mathbb{E}[X] = \frac{a+b}{2}, \quad \mathrm{Var}(X) = \frac{(b-a+1)^2 - 1}{12}.$$

6 Continuous Random Variables

Uniform(a, b)

$f(x) = \frac{1}{b-a}$ on (a, b).

$$\mathbb{E}[X] = \frac{a+b}{2}, \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

Exponential(λ)

$f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.

$$\mathbb{E}[X] = \frac{1}{\lambda}, \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Memoryless: $\mathbb{P}(X > t + s \mid X > s) = e^{-\lambda t}$.

Gaussian / Normal $\mathcal{N}(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \mathbb{E}[X] = \mu, \ \mathrm{Var}(X) = \sigma^2.$$

If $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ independent, then $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \ \sigma_1^2 + \sigma_2^2)$. Standardization: $Z = (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$.

Gamma(α, β) **(shape** α**, rate** β**)**

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \ x \geq 0; \quad \mathbb{E}[X] = \frac{\alpha}{\beta}, \ \mathrm{Var}(X) = \frac{\alpha}{\beta^2}.$$

Sum of independent Gamma(α_i, β) with common rate is Gamma($\sum \alpha_i, \beta$). Exponential is Gamma($1, \lambda$).

Beta(a, b) **on** $(0, 1)$

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad \mathbb{E}[X] = \frac{a}{a+b}, \ \mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Conjugate prior for Bernoulli/Binomial.

7 Joint Distributions and Independence of r.v.s

For discrete (X, Y): $p_{X,Y}(x, y)$; continuous: $f_{X,Y}(x, y)$. Marginals: $p_X(x) = \sum_y p_{X,Y}(x, y)$ or $f_X(x) = \int f_{X,Y}(x, y)\, dy$.

$$X \perp Y \iff p_{X,Y}(x, y) = p_X(x)p_Y(y) \text{ (discrete)} \quad \text{or} \quad f_{X,Y}(x, y) = f_X(x)f_Y(y) \text{ (continuous)}.$$

Conditional distributions: $p_{X|Y}(x \mid y) = \dfrac{p_{X,Y}(x, y)}{p_Y(y)}$, $f_{X|Y}(x \mid y) = \dfrac{f_{X,Y}(x, y)}{f_Y(y)}$ (when denominators > 0).

8 Transforms and Common Tools (Very Short)

Characteristic function $\varphi_X(t) = \mathbb{E}[e^{itX}]$ always exists; independence $\Rightarrow$ product of characteristic functions. CLT: sums of i.i.d. (finite variance) approximate normal. Chebyshev/Markov inequalities bound tails via moments.

Exercises

1. **Geometric distribution properties** Let $X \sim \text{Geom}(p)$ be independent geometric random variables (support $\{1, 2, \dots\}$, representing the trial of the first success). Prove the *memoryless property*: for any integers $m, n \geq 0$,

$$\mathbb{P}(X > m + n \mid X > m) = \mathbb{P}(X > n).$$

2. **Exponential distribution properties** Let $X \sim \text{Exp}(\lambda)$ be independent exponential random variables. Prove the *memoryless property*: for any $s, t \geq 0$,

$$\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t).$$

3. **Closure of the Gaussian distribution** Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ be independent.

   (a) Show that $X + Y \sim \mathcal{N}(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2)$.

   (b) Generalize the result to the sum of n independent Gaussian random variables.

4. **Relationship between exponential distribution and Poisson process**

   (a) Let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ. Show that $N(t) \sim \text{Poisson}(\lambda t)$.

   (b) Prove that the interarrival time between two consecutive events follows an exponential distribution $\text{Exp}(\lambda)$.

   (c) Prove that interarrival times are i.i.d., establishing the link between the exponential distribution and the Poisson process.

Proof Sketches

1. Geometric distribution

(a) $X_1 + X_2$ is the number of trials needed for two successes. For $k \geq 2$:

$$\mathbb{P}(X_1 + X_2 = k) = \binom{k-1}{1}(1-p)^{k-2}p^2,$$

which is a negative binomial $NB(r = 2, p)$. In general, the sum of n i.i.d. geometric r.v.s is $NB(r = n, p)$.

(b) For $X \sim \text{Geom}(p)$:

$$\mathbb{P}(X > m + n \mid X > m) = \frac{(1-p)^{m+n}}{(1-p)^m} = (1-p)^n = \mathbb{P}(X > n).$$

2. Exponential distribution

(a) By convolution:

$$f_{X_1+X_2}(s) = \int_0^s \lambda e^{-\lambda x}\, \lambda e^{-\lambda(s-x)}\, dx = \lambda^2 s e^{-\lambda s}, \quad s \geq 0,$$

which is $\text{Gamma}(2, \lambda)$.

(b) Memorylessness:

$$\mathbb{P}(X > t + s \mid X > s) = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(X > t).$$

3. Gaussian distribution

Characteristic function method:

$$\varphi_X(t) = \exp\!\left(i\mu_1 t - \tfrac{1}{2}\sigma_1^2 t^2\right), \quad \varphi_Y(t) = \exp\!\left(i\mu_2 t - \tfrac{1}{2}\sigma_2^2 t^2\right).$$

Thus

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = \exp\!\left(i(\mu_1 + \mu_2)t - \tfrac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2\right).$$

Hence $X + Y \sim \mathcal{N}(\mu_1 + \mu_2,\, \sigma_1^2 + \sigma_2^2)$. The generalization to n independent Gaussians follows similarly.

4. Exponential distribution and Poisson process

(a) By definition of Poisson process: $N(t) \sim \text{Poisson}(\lambda t)$.

(b) For the first interarrival time T_1:

$$\mathbb{P}(T_1 > t) = \mathbb{P}(N(t) = 0) = e^{-\lambda t}, \quad t \geq 0,$$

so $T_1 \sim \text{Exp}(\lambda)$.

(c) By the independent increments property, all interarrival times are i.i.d. exponential with rate λ. Thus the Poisson counting process and the exponential waiting-time distribution are two sides of the same structure.