MA2116T (AY25/26)
Probability (T)
by Prof Sun Rongfeng
Compiled by ygh3rn

Def. 1 (Axioms of Probability)
Let Ω be the sample space. A probability measure $\mathbb{P}$ must satisfy:
1. For any event $E \subset \Omega$, $\mathbb{P}(E) \in [0,1]$.
2. $\mathbb{P}(\Omega) = 1$.
3. **(Countable Additivity)** For any sequence of pairwise disjoint events $(E_n)_{n\in\mathbb{N}}$,
$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(E_n).$$

Rem. 2
The *null event* $\emptyset$ satisfies $\mathbb{P}(\emptyset) = 0$. The *sure event* Ω satisfies $\mathbb{P}(\Omega) = 1$.

Prop. 3
Let P be a probability measure on a sample space Ω.
1. **(Finite additivity)** If $E_1, \dots, E_n$ are mutually exclusive, then $\mathbb{P}(E_1 \cup \cdots \cup E_n) = \sum_{i=1}^n \mathbb{P}(E_i)$.
2. $\mathbb{P}(\emptyset) = 0$.
3. $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$.
4. For any event A, $\mathbb{P}(A) = \mathbb{P}(A \cap E) + \mathbb{P}(A \cap E^c)$.
5. If two events E, F satisfy $E \subset F$, then $\mathbb{P}(E) \le \mathbb{P}(F)$.

Thm. 4 (Inclusion–Exclusion Principle)
Given events $E_1, E_2, \dots, E_n$,
$$\mathbb{P}\Big(\bigcup_{i=1}^{n} E_i\Big) = \sum_{\emptyset \neq J \subset \{1,\dots,n\}} (-1)^{|J|+1}\, \mathbb{P}\Big(\bigcap_{j\in J} E_j\Big).$$

Thm. 5 (Boole's Inequality)
Given countable index set I and events $(E_i)_{i\in I}$,
$$\mathbb{P}\left(\bigcup_{i\in I} E_i\right) \le \sum_{i\in I} \mathbb{P}(E_i).$$

Prop. 6 (Derangements)
For $N \ge 1$, the probability a random permutation has no fixed points is
$$\mathbb{P}_N = \sum_{k=0}^{N} \frac{(-1)^k}{k!}.$$

Def. 7 (Conditional Probability)
Let Ω be a sample space and $\mathbb{P}$ a probability measure on Ω. Given two events $A, B \subset \Omega$, the *conditional probability* of the event B given A is defined by
$$\mathbb{P}(B \mid A) := \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}.$$

Def. 8 (Mult. Rule for Successive Conditioning)
Given events $E_1, E_2, \dots, E_n$,
$$\mathbb{P}(E_1 \cap \cdots \cap E_n)$$
$$= \mathbb{P}(E_1)\,\mathbb{P}(E_2 \mid E_1) \cdots \mathbb{P}(E_n \mid E_1 \cap \cdots \cap E_{n-1}),$$
provided that $\mathbb{P}(E_1), \mathbb{P}(E_1 \cap E_2), \dots, \mathbb{P}(E_1 \cap \cdots \cap E_n) > 0$.

Thm. 9 (Law of Total Probability)
Let $E_1, E_2, \dots, E_n$ be a partition of the sample space Ω. Then for any event A,
$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(E_i)\mathbb{P}(A \mid E_i).$$

Thm. 10 (Bayes' Rule)
Let $E_1, E_2, \dots, E_n$ be a partition of Ω and A any event with $\mathbb{P}(A) > 0$. Then for each $1 \le k \le n$,
$$\mathbb{P}(E_k \mid A) = \frac{\mathbb{P}(E_k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(E_k)\mathbb{P}(A \mid E_k)}{\sum_{i=1}^n \mathbb{P}(E_i)\mathbb{P}(A \mid E_i)}.$$

Def. 11 (Independence of Two Events)
Two events A and B are said to be *independent* if
$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Prop. 12
Let A, B be two events.
1. If A and B are independent with $\mathbb{P}(A), \mathbb{P}(B) > 0$, then $\mathbb{P}(A \mid B) = \mathbb{P}(A)$, $\quad \mathbb{P}(B \mid A) = \mathbb{P}(B)$.
2. Suppose that $\mathbb{P}(A), \mathbb{P}(B) > 0$. If A and B are independent, then $A \cap B \neq \emptyset$.
3. The sure event Ω and the null event $\emptyset$ is independent of all other events.

Def. 13 (Joint Independence)
A collection of events $E_1, \dots, E_n$ are said to be *jointly independent* if for any non-empty index set $I \subset \{1, \dots, n\}$,
$$\mathbb{P}\Big(\bigcap_{i\in I} E_i\Big) = \prod_{i\in I} \mathbb{P}(E_i).$$

Rem. 14
The above equation contains $2^n - 1$ different cases.
Pairwise independence among $E_1, \dots, E_n$ does not imply their joint independence.

Prop. 15 (Gambler's Ruin)
If $p = \frac{1}{2}$, then for $0 \le k \le N$,
$$w(k) = \frac{k}{N}.$$
If $p \neq \frac{1}{2}$ and $r = \frac{1-p}{p}$, then for $1 \le n \le N-1$,
$$w(n) = \frac{1 - r^n}{1 - r^N}.$$
If A starts with n and B with $N - n$, then B's winning probability is
$$u(N-n) = \frac{r^n - r^N}{1 - r^N}.$$

Def. 16 (Cumulative Distribution Function)
The *cumulative distribution function* of a r.v. X is the function $F_X : \mathbb{R} \to [0,1]$ defined by
$$F_X(x) := P_X((-\infty, x]) = \mathbb{P}(X \le x), \quad \text{for all } x \in \mathbb{R}.$$

Rem. 17
This amounts to specifying $P_X(A)$ for all $A \subset \mathbb{R}$ of the form $A = (-\infty, x]$. If X is a d.r.v., then its c.d.f. F_X is a *step function*. The *jumps* occur at the points in *Range*(X), with jump size
$$F_X(x) - F_X(x^-) = \mathbb{P}(X = x).$$

Prop. 18 (Properties of the c.d.f.)
Let F_X be the c.d.f. of a real-valued r.v. X. Then:
1. $0 \le F_X(x) \le 1$ for all $x \in \mathbb{R}$.
2. $F_X(a) \le F_X(b)$ for all $a < b$. More precisely,
$$F_X(b) - F_X(a) = \mathbb{P}(X \in (a, b]) \ge 0.$$
3. $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$ (by countable additivity).
4. F_X is right-continuous, i.e., $\lim_{\varepsilon\downarrow 0} F_X(x+\varepsilon) = F_X(x)$ for all x.
5. F_X has left-hand limits, i.e., $F_X(x^-) := \lim_{\varepsilon\downarrow 0} F_X(x - \varepsilon)$ exists.

Def. 19 (Probability Mass Function)
The *probability mass function* of a d.r.v. X is the function $f_X : \mathbb{R} \to [0,1]$ defined by
$$f_X(x) := \begin{cases} \mathbb{P}(X = x), & x \in \mathit{Range}(X), \\ 0, & \text{otherwise.} \end{cases}$$

Prop. 20 (Properties of the p.m.f.)
Let f_X be the p.m.f. of a d.r.v. X. Then:
1. $f_X(x) \ge 0$ for all x.
2. $\sum_x f_X(x) = 1$.
3. $\mathbb{P}(X \in A) = \sum_{x\in A} f_X(x)$ for any $A \subset \mathbb{R}$.
4. $F_X(x) = \sum_{t \le x} f_X(t)$ for all $x \in \mathbb{R}$.
5. $f_X(x) = F_X(x) - F_X(x^-)$ for all $x \in \mathbb{R}$.

Def. 21 (Continuous Random Variable)
A r.v. X is called *continuous* if $\mathbb{P}(X = x) = 0$ for every $x \in \mathbb{R}$.

Rem. 22
Since $\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$, X being continuous is equivalent to $F_X(x)$ being continuous at every $x \in \mathbb{R}$.

Def. 23 (Absolutely Continuous r.v.)
A r.v. X is called *absolutely continuous* if it admits a probability density function $f_X : \mathbb{R} \to [0, \infty)$ such that
1. $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$;
2. $\mathbb{P}(X \in (a, b]) = \int_a^b f_X(x)\,dx$.

Prop. 24 (Properties of the p.d.f.)
For an absolutely c.r.v. X, f_X also satisfies:
1. For $A \subseteq \mathbb{R}$, $\mathbb{P}(X \in A) = \int_A f_X(x)\,dx$;
2. $F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$;
3. $f_X(x) = \frac{d}{dx}F_X(x) = F_X'(x)$ if the derivative $F_X'(x)$ exists at x;
4. For $a < b$,
$$\mathbb{P}(X \in (a,b)) = \mathbb{P}(X \in [a,b]) = F_X(b) - F_X(a).$$

Def. 25 (Expectation of d.r.v.)
If X is a d.r.v. taking values in the set $E = \{x_i : i \in \mathbb{N}\} \subset \mathbb{R}$ and has p.m.f. f_X, then its *expectation/mean* is defined by
$$\mathbb{E}[X] := \sum_i x_i f_X(x_i).$$

Def. 26 (Expectation of c.r.v.)
If X is a continous real-valued r.v. with p.d.f. f_X, then its *expectation/mean* is defined by
$$\mathbb{E}[X] := \int_{\mathbb{R}} x f_X(x)\,dx.$$

Def. 27 (Expectation of Functions of r.v.)
Let X be a real-valued r.v.. Given a function $g : \mathbb{R} \to \mathbb{R}$, define $Y := g(X)$, which is also a r.v.. Its *expectation* is defined as
$$\mathbb{E}[g(X)] = \begin{cases} \sum_x g(x) f_X(x), & \text{if } X \text{ is discrete,} \\ \int_{\mathbb{R}} g(x) f_X(x)\,dx, & \text{if } X \text{ is a.c..} \end{cases}$$

Prop. 28 (Properties of Expectation)
For any functions $g, h : \mathbb{R} \to \mathbb{R}$ and $a, b \in \mathbb{R}$,
1. **(Comparison)**
$$\mathbb{P}(X \ge a) = 1 \;\Rightarrow\; \mathbb{E}[X] \ge a,$$
$$\mathbb{P}(X \le b) = 1 \;\Rightarrow\; \mathbb{E}[X] \le b.$$
2. **(Linearity)**
$$\mathbb{E}[ag(X) + bh(X)] = a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)].$$

Def. 29 (Variance)
The *variance* of a r.v. X is defined as
$$\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$
The *standard deviation* of X is defined as
$$\sigma_X := \sqrt{\mathrm{Var}(X)}.$$

Rem. 30
Since $\mathrm{Var}(X) \ge 0$, we have
$$\mathbb{E}[X^2] = (\mathbb{E}[X])^2 + \mathrm{Var}(X) \;\ge\; (\mathbb{E}[X])^2.$$
For $a, b \in \mathbb{R}$,
$$\mathrm{Var}(X + b) = \mathrm{Var}(X),$$
$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X).$$

Prop. 31 (Markov's Inequality)
Let $X \ge 0$ be a r.v. and $c > 0$. Then
$$\mathbb{P}(X \ge c) \le \frac{\mathbb{E}[X]}{c}.$$

Prop. 32 (Chebyshev's Inequality)
Let X be a r.v. with $\mathbb{E}[X], \mathrm{Var}(X) < \infty$. For any $k > 0$,
$$\mathbb{P}\big(|X - \mathbb{E}[X]| \ge k\big) \le \frac{\mathrm{Var}(X)}{k^2}.$$

Prop. 33 (Jensen's Inequality)
Let g be a convex (concave) function and X be a r.v. such that $\mathbb{E}[|X|] < \infty$. Then
$$g(\mathbb{E}[X]) \le (\ge)\; \mathbb{E}[g(X)].$$
Equality holds iff g is linear or X is a.s. constant.

Prop. 34 (Tail-Sum Formula)
Let X be a r.v. taking values in $\{0, 1, 2, \dots\}$. Then
$$\mathbb{E}[X] = \sum_{i=1}^{\infty} i\,\mathbb{P}(X = i) = \sum_{n=1}^{\infty} \mathbb{P}(X \ge n),$$
i.e., as the sum of X's *tail probabilities.*

Prop. 35 (Tail-Integral Formula)
$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X \ge t)\,dt,$$
and for $k > 0$,
$$\mathbb{E}[X^k] = \int_0^{\infty} kt^{k-1}\mathbb{P}(X \ge t)\,dt.$$

Def. 36 (Bernoulli Distribution)
A r.v. X with
$$\mathbb{P}(X = 1) = p \quad \text{and} \quad \mathbb{P}(X = 0) = 1 - p$$
for some $p \in [0,1]$, is called a *Bernoulli random variable* with parameter p, denoted by $X \sim \mathrm{Ber}(p)$. Its probability distribution is called the *Bernoulli distribution* with parameter p.

Def. 37 (Indicator Random Variable)
Let $A \subset \Omega$ be an event. Define $X : \Omega \to \{0, 1\}$ by
$$X(\omega) := 1_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \in A^c. \end{cases}$$
Then $X \sim \mathrm{Ber}(p)$ with $p = \mathbb{P}(A)$, and $\mathbb{E}[1_A] = p$. X is called the *indicator random variable* for the event A.

Def. 38 (Binomial Distribution)
Let $X_1, X_2, \dots, X_n$ be independent $\mathrm{Ber}(p)$ r.v.'s. The number of successes among the first n Bernoulli trials,
$$Y := X_1 + \cdots + X_n,$$
is called a *Binomial random variable* with parameters n and p, denoted by $Y \sim \mathrm{Bin}(n,p)$. Its p.m.f. is
$$\mathbb{P}(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 \le k \le n.$$

Prop. 39
If $Y \sim \mathrm{Bin}(n,p)$, then
$$\mathbb{E}[Y] = np, \quad \mathrm{Var}(Y) = np(1-p).$$

Prop. 40
Let $X_1, \dots, X_n$ be independent r.v.'s. Then
$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i], \quad \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Def. 41 (Poisson Distribution)
A r.v. X taking values in $\mathbb{N}_0 := \{0, 1, 2, \dots\}$ is said to have a *Poisson distribution* with parameter $\lambda > 0$, denoted $X \sim \mathrm{Pois}(\lambda)$, if
$$\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k \in \mathbb{N}_0.$$

Prop. 42
If $X \sim \mathrm{Pois}(\lambda)$, then
$$\mathbb{E}[X] = \lambda, \quad \mathrm{Var}(X) = \lambda.$$

Thm. 43 (Poisson Limit Theorem)
Let $\lambda > 0$. For $n \in \mathbb{N}$, let $X_{n,1}, \dots, X_{n,n}$ be i.i.d. $\mathrm{Ber}(\lambda/n)$ r.v.'s. Define
$$Y_n := \sum_{i=1}^{n} X_{n,i} \sim \mathrm{Bin}\big(n, \tfrac{\lambda}{n}\big), \quad Z \sim \mathrm{Pois}(\lambda).$$
Then
$$\lim_{n\to\infty} \mathbb{P}(Y_n = k) = \mathbb{P}(Z = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad \forall k \in \mathbb{N}_0.$$
Hence $Y_n \xrightarrow{d} Z$.

Def. 44 (Discrete Uniform Distribution)
Let $A = \{x_1, x_2, \dots, x_k\}$ with $x_1 < x_2 < \cdots < x_k$. A random variable X is said to have the *discrete uniform distribution* on A if
$$\mathbb{P}(X = x_i) = \frac{1}{k}, \quad 1 \le i \le k.$$

Prop. 45
If X has the discrete uniform distribution on A, then
$$\mathbb{E}[X] = \frac{1}{k}\sum_{i=1}^{k} x_i, \quad \mathrm{Var}(X) = \frac{1}{k}\sum_{i=1}^{k} x_i^2 - \left(\frac{1}{k}\sum_{i=1}^{k} x_i\right)^2.$$

Def. 46 (Geometric Distribution)
A random variable X taking values in $\mathbb{N}$ is called a *geometric random variable* with parameter $p \in (0,1)$, denoted $X \sim \mathrm{Geom}(p)$, if
$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k \in \mathbb{N}.$$
Its distribution is called the *geometric distribution with parameter* p, because its p.m.f. is a geometric sequence, with $\sum_k \mathbb{P}(X = k) = 1$.

Prop. 47 (Tail Probability Formula)
For $X \sim \mathrm{Geom}(p)$, the tail probability has the form
$$\mathbb{P}(X \ge k) = \sum_{i=k}^{\infty} p(1-p)^{i-1} = (1-p)^{k-1}.$$

Prop. 48 (Memorylessness)
$X \sim \mathrm{Geom}(p)$ is *memoryless* in the sense that, for any $k \in \mathbb{N}$,
$$\mathbb{P}(X - k = i \mid X > k) = \mathbb{P}(X = i), \quad \forall i \in \mathbb{N}.$$

Prop. 49
If $X \sim \mathrm{Geom}(p)$, then
$$\mathbb{E}[X] = \frac{1}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2}.$$

Def. 50 (Negative Binomial Distribution)
Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. Bernoulli trials with success parameter $p \in (0,1)$. Fix $r \in \mathbb{N}$, and let X be the number of trials needed to see the r-th success. Then X is called a *negative binomial random variable* with parameters (r, p), denoted
$$X \sim \mathrm{NB}(r, p),$$
with p.m.f.
$$\mathbb{P}(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n \ge r.$$

Rem. 51
This distribution is called "negative binomial" because of the negative binomial expansion
$$(1-t)^{-r} = \sum_{n=r}^{\infty} \binom{n-1}{r-1} t^{n-r}, \quad |t| < 1,$$
which shows that $\sum_n \mathbb{P}(X = n) = 1$.

Prop. 52
If $X \sim \mathrm{NB}(r, p)$, then
$$\mathbb{E}[X] = \frac{r}{p}, \quad \mathrm{Var}(X) = \frac{r(1-p)}{p^2}.$$

Prop. 53
Suppose $X \sim \mathrm{NB}(r, p)$ and $Y \sim \mathrm{NB}(r+1, p)$. Then
$$\mathbb{E}[X^k] = \frac{r}{p}\mathbb{E}\big[(Y-1)^{k-1}\big], \quad k \in \mathbb{N}.$$

Def. 54 (Hypergeometric Distribution)
Sample n balls without replacement from an urn containing N balls, of which m are white and $N - m$ are black. Let X be the number of white balls selected. Then X is called a *hypergeometric random variable* with parameters (n, N, m), denoted
$$X \sim H(n, N, m),$$
with p.m.f.
$$\mathbb{P}(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}, \quad 0 \le i \le n,$$
assuming both colors have more than n balls.

Prop. 55
If $X \sim H(n, N, m)$, then
$$\mathbb{E}[X] = n \cdot \frac{m}{N}, \quad \mathrm{Var}(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{m}{N}\left(1 - \frac{m}{N}\right).$$

Prop. 56
Let $X \sim H(n, N, m)$ and $Y \sim H(n-1, N-1, m-1)$. Then
$$\mathbb{E}[X^k] = \frac{nm}{N} E\big[(Y+1)^{k-1}\big], \quad k \in \mathbb{N}.$$

Def. 57 (Uniform Distribution)
Given $a < b$, a c.r.v. X is said to have a *uniform distribution* over (a, b), denoted
$$X \sim U(a, b),$$
if it has p.d.f.
$$f_X(x) = \frac{\mathbf{1}_{[a,b]}(x)}{b-a} = \begin{cases} \frac{1}{b-a}, & x \in (a,b), \\ 0, & x \notin (a,b). \end{cases}$$

Prop. 58
If $X \sim U(a, b)$, then
$$\mathbb{E}[X] = \frac{a+b}{2}, \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

Rem. 59
For any interval $(x, y) \subset (a, b)$,
$$\mathbb{P}(x < X < y) = \frac{y-x}{b-a}.$$

Def. 60 (Exponential Distribution)
A c.r.v. X is said to have an *exponential distribution* with parameter $\lambda > 0$, denoted
$$X \sim \mathrm{Exp}(\lambda),$$
if it has p.d.f.
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

Prop. 61
If $X \sim \mathrm{Exp}(\lambda)$, then
$$\mathbb{E}[X] = \frac{1}{\lambda}, \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Prop. 62
If $X \sim \mathrm{Exp}(\lambda)$ and $Y \sim \mathrm{Exp}(1)$, then
$$X \overset{dist}{=} \frac{Y}{\lambda}.$$

Prop. 63 (Memorylessness)
If $X \sim \mathrm{Exp}(\lambda)$, then X is *memoryless* in the sense that, for any $s, t \ge 0$,
$$\mathbb{P}(X - t \ge s \mid X \ge t) = \mathbb{P}(X \ge s).$$

Rem. 64
The exponential distribution is the only family of continuous distributions that satisfies the memorylessness property.

Thm. 65 (Geometric → Exponential)
Fix $\lambda > 0$ and let $\delta \in (0, 1/\lambda)$. Let $X_\delta \sim \mathrm{Geom}(\lambda\delta)$. Then as $\delta \downarrow 0$, the sequence of r.v.'s δX_δ converges in distribution to $Y \sim \mathrm{Exp}(\lambda)$, in the sense that
$$\lim_{\delta\downarrow 0} \mathbb{P}(\delta X_\delta > t) = e^{-\lambda t} = \mathbb{P}(Y > t), \quad \forall t > 0.$$

Def. 66 (Gamma Distribution)
Let $\alpha > 0$ and $\lambda > 0$. A c.r.v. Y is said to have a *Gamma distribution* with parameters (α, λ), denoted $Y \sim \Gamma(\alpha, \lambda)$, if it has p.d.f.
$$f_Y(y) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y}, & y > 0, \\ 0, & y \le 0. \end{cases}$$
where the Gamma function is defined by
$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x}\,dx, \quad \alpha > 0,$$
and satisfies the recurrence
$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha), \quad \Gamma(1) = 1.$$
In particular, for $n \in \mathbb{N}$, $\Gamma(n+1) = n!$.

Rem. 67
If $\alpha \in \mathbb{N}$ and $X_1, \dots, X_\alpha \sim \mathrm{Exp}(\lambda) = \Gamma(1, \lambda)$ i.i.d., then
$$Y = X_1 + \cdots + X_\alpha \sim \Gamma(\alpha, \lambda),$$
i.e. the waiting time for the α-th event to occur.

Prop. 68
If $Y \sim \Gamma(\alpha, \lambda)$ with $\alpha \in \mathbb{N}$, then since $Y = X_1 + \cdots + X_\alpha$ with $X_i \sim \mathrm{Exp}(\lambda)$ i.i.d.,
$$\mathbb{E}[Y] = \frac{\alpha}{\lambda}, \quad \mathrm{Var}(Y) = \frac{\alpha}{\lambda^2}.$$
These identities extend to all $\alpha > 0$.

Def. 69 (Normal / Gaussian Distribution)
A c.r.v. X is said to have a *normal distribution* with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, denoted $X \sim N(\mu, \sigma^2)$, if it has p.d.f.
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$
When $\mu = 0$, $\sigma^2 = 1$, X is called a *standard normal* random variable.

Rem. 70
Using $x = y^2/2$, one finds
$$\int_0^{\infty} e^{-y^2/2}\,dy = \frac{1}{\sqrt{2}}\Gamma\big(\tfrac{1}{2}\big).$$
Thus
$$\Gamma\big(\tfrac{1}{2}\big) = \sqrt{\pi}.$$

Prop. 71
If $X \sim N(\mu, \sigma^2)$, then
$$\mathbb{E}[X] = \mu, \quad \mathrm{Var}(X) = \sigma^2.$$

Thm. 72 (De Moivre–Laplace Theorem)
Let $X \sim \mathrm{Bin}(n, p)$ with $p \in (0,1)$. Then for any $a < b$,
$$\mathbb{P}\left(\frac{X - np}{\sqrt{np(1-p)}} \in (a, b)\right) \xrightarrow[n\to\infty]{} \mathbb{P}(Z \in (a, b)),$$
where $Z \sim N(0, 1)$.

Rem. 73 (Continuity Correction)
Since $X \sim B(n, p)$ is integer-valued, while its normal approximation $Y \sim N(np, np(1-p))$ is continuous, one applies a continuity correction:
$$\mathbb{P}(X = x) = \mathbb{P}\big(X \in (x - \tfrac{1}{2}, x + \tfrac{1}{2})\big)$$
$$\approx \mathbb{P}\big(Y \in (x - \tfrac{1}{2}, x + \tfrac{1}{2})\big).$$

Thm. 74 (Affine Transformation of Normal r.v.)
If $X \sim N(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$, then
$$Y := aX + b \;\sim\; N(a\mu + b,\, a^2\sigma^2).$$
$$\mathbb{E}[Y] = a\mu + b, \quad \mathrm{Var}(Y) = a^2\sigma^2.$$

Prop. 75
If $X \sim N(\mu, \sigma^2)$, then
$$X = \sigma Z + \mu,$$
where $Z \sim N(0, 1)$ is a standard normal r.v..

Prop. 76 (Leibniz Integral Rule)
Let g be an integrable function, and let $\varphi : \mathbb{R} \to \mathbb{R}$ be differentiable. Define
$$F(y) := \int_a^{\varphi(y)} g(x)\,dx.$$
Then
$$F'(y) = g(\varphi(y))\,\varphi'(y).$$

Thm. 77 (Function of a r.v.)
Let X be a c.r.v. with p.d.f. f_X, and let $g : \mathbb{R} \to \mathbb{R}$ be a differentiable and monotone function. Define $Y := g(X)$. Then the p.d.f. of Y is
$$f_Y(y) = f_X(g^{-1}(y))\,\frac{d}{dy}g^{-1}(y).$$

Def. 78 (Joint Probability Mass Function)
The *joint probability mass function* of two d.r.v.'s X and Y defined on the same sample space is given by
$$f_{X,Y}(x, y) := \mathbb{P}(X = x, Y = y)$$
$$= \mathbb{P}\big(\{\omega \in \Omega : X(\omega) = x,\ Y(\omega) = y\}\big).$$
For three or more d.r.v.'s, the definition is analogous.

Prop. 79 (Properties of the Joint p.m.f.)
Let (X, Y) be two d.r.v.'s with joint p.m.f. $f_{X,Y}$. Then:
1. $f_{X,Y}(x, y) \ge 0$ for all x, y.
2. It is a p.m.f., i.e.
$$\sum_x \sum_y f_{X,Y}(x, y) = 1.$$
3. For any $A \subseteq \mathbb{R}^2$ (assuming X, Y are real-valued),
$$\mathbb{P}\big((X, Y) \in A\big) = \sum_{(x,y)\in A} f_{X,Y}(x, y).$$
4. The *marginal p.m.f.* of X is
$$f_X(x) = \mathbb{P}(X = x) = \sum_y f_{X,Y}(x, y).$$

Def. 80 (Joint Probability Density Function)
Let (X,Y) be a random vector taking values in $\mathbb{R}^2$. We say (X,Y) admits a *joint probability density function* $f_{X,Y}$ if:
1. $f_{X,Y}(x,y)\ge 0$ for all $(x,y)\in\mathbb{R}^2$.
2. $\int_{\mathbb{R}}\int_{\mathbb{R}} f_{X,Y}(x,y)\,dx\,dy = 1$.
3. For any $A\subseteq\mathbb{R}^2$,
$$\mathbb{P}\big((X,Y)\in A\big) = \iint_A f_{X,Y}(x,y)\,dx\,dy.$$

Prop. 81 (Marginal Density)
If (X,Y) has joint p.d.f. $f_{X,Y}$, then the *marginal density* f_X is given by
$$f_X(x)=\int_{\mathbb{R}} f_{X,Y}(x,y)\,dy.$$

Def. 82 (Joint c.d.f.)
The *joint cumulative distribution function* of (X,Y) is
$$F_{X,Y}(x,y) := \mathbb{P}(X\le x,\, Y\le y).$$
The marginal density of X can be recovered as
$$F_X(x)=\lim_{y\to\infty}F_{X,Y}(x,y).$$

Prop. 83
If (X,Y) is discrete:
$$F_{X,Y}(x,y)=\sum_{a\le x,\, b\le y} f_{X,Y}(a,b),$$
$$f_{X,Y}(x,y) = \mathbb{P}(X\le x, Y\le y) - \mathbb{P}(X\le x^-, Y\le y) - \mathbb{P}(X\le x, Y\le y^-) + \mathbb{P}(X\le x^-, Y\le y^-).$$
If (X,Y) is continuous:
$$F_{X,Y}(x,y)=\int_{-\infty}^x\int_{-\infty}^y f_{X,Y}(a,b)\,db\,da,$$
$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x\,\partial y}F_{X,Y}(x,y).$$

Prop. 84 (Cauchy–Schwarz Inequality)
If X and Y are real-valued r.v.'s with $\mathbb{E}[X^2],\mathbb{E}[Y^2]<\infty$, then
$$\big|\mathbb{E}[XY]\big| \le \mathbb{E}[X^2]^{1/2}\,\mathbb{E}[Y^2]^{1/2}.$$

Prop. 85 (Paley–Zygmund Inequality)
Let Z be a non-negative r.v. with $\mathbb{E}[Z^2]<\infty$. Then for any $a\in[0,1]$,
$$\mathbb{P}(Z\ge a\,\mathbb{E}[Z]) \ \ge\ (1-a)^2\,\frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]}.$$

Def. 86 (Conditional Expectation and Variance)
Let X be a real-valued r.v. and $A\in\mathcal{F}$ with $\mathbb{P}(A)>0$. The *conditional expectation* of X given A is
$$\mathbb{E}[X\mid A] := \frac{\mathbb{E}[X\mathbf{1}_A]}{\mathbb{P}(A)}.$$
The *conditional variance* of X given A is
$$\operatorname{Var}(X\mid A) := \mathbb{E}[X^2\mid A] - \big(\mathbb{E}[X\mid A]\big)^2 = \mathbb{E}\Big[(X-\mathbb{E}[X\mid A])^2\,\Big|\,A\Big].$$

Def. 87 (Conditioning on a d.r.v.)
Let X,Y be r.v.'s with Y taking values in a countable set S. For $y\in S$ with $\mathbb{P}(Y=y)>0$, the conditional distribution of X given $Y=y$ is
$$\mathbb{P}(X\in\cdot\mid Y=y),$$
with conditional expectation $\mathbb{E}[X\mid Y=y]$.

Prop. 88 (Law of Total Expectation)
For r.v.'s X,Y and any measurable function $f:\mathbb{R}\to\mathbb{R}$,
$$\mathbb{E}[f(X)] = \mathbb{E}[\mathbb{E}[f(X)\mid Y]].$$
In particular, for any $A\subseteq\mathbb{R}$,
$$\mathbb{P}(X\in A)=\mathbb{E}[\mathbb{P}(X\in A\mid Y)].$$

Prop. 89 (Conditional Density)
Let (X,Y) be jointly absolutely continuous with joint density $f_{X,Y}$. Then the conditional density of X given $Y=y_0$ is
$$f_{X|Y}(x\mid Y=y_0) = \frac{f_{X,Y}(x,y_0)}{f_Y(y_0)}.$$

Prop. 90
For jointly absolutely continuous (X,Y),
$$f_{X,Y}(x,y) = f_Y(y)f_{X|Y}(x\mid y) = f_X(x)f_{Y|X}(y\mid x).$$

Prop. 91
For any measurable function $g:\mathbb{R}^2\to\mathbb{R}$,
$$\mathbb{E}[g(X,Y)] = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} g(x,y)f_{X|Y}(x\mid y)\,dx\right) f_Y(y)\,dy.$$

Prop. 92 (Tower Property)
For r.v.'s $X_1,\dots,X_n$ and any measurable φ,
$$\mathbb{E}[\varphi(X_1,\dots,X_n)] = \mathbb{E}\Big[\cdots\mathbb{E}\big[\mathbb{E}[\varphi(X_1,\dots,X_n)\mid X_1,\dots,X_{n-1}]\cdots\mid X_1\big]\Big].$$

Prop. 93 (Law of Total Variance)
Let X,Y be r.v.'s with finite $\mathbb{E}[X^2]$. Then
$$\operatorname{Var}(X) = \mathbb{E}\big[\operatorname{Var}(X\mid Y)\big] + \operatorname{Var}\big(\mathbb{E}[X\mid Y]\big).$$

Def. 94 (Independent Random Variables)
Two r.v.'s X and Y are said to be *independent* if
$$\mathbb{P}(X\in A, Y\in B) = \mathbb{P}(X\in A)\,\mathbb{P}(Y\in B)$$
for any $A,B\subset\mathbb{R}$.
More generally, a collection of r.v.'s $X_1,X_2,\dots,X_n$ are *jointly independent* if for any measurable events $A_1,\dots,A_n$,
$$\mathbb{P}(X_i\in A_i \text{ for all } 1\le i\le n) = \prod_{i=1}^n \mathbb{P}(X_i\in A_i).$$
If $\mathbf{X} := (X_1,\dots,X_n)$ has joint p.m.f. (or p.d.f.) $f_{\mathbf{X}}(x_1,\dots,x_n)$, then their joint independence is equivalent to
$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n f_{X_i}(x_i),\quad \mathbf{x}\in\mathbb{R}^n.$$

Prop. 95 (Expectation of a Function of r.v.)
Let X and Y be two r.v.'s, and let $g:\mathbb{R}^2\to\mathbb{R}$. Then
$$\mathbb{E}[g(X,Y)] = \begin{cases} \displaystyle\sum_{(x,y)} g(x,y)f_{X,Y}(x,y), & \text{(d.r.v.'s)},\\ \displaystyle\iint g(x,y)f_{X,Y}(x,y)\,dx\,dy, & \text{(c.r.v.'s)}.\end{cases}$$

Prop. 96
If $X_1,\dots,X_n$ are jointly independent r.v.'s, and $\varphi_1,\dots,\varphi_n$ are real-valued functions, then
$$\mathbb{E}\left[\prod_{i=1}^n\varphi_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}[\varphi_i(X_i)].$$

Def. 97 (Covariance)
The covariance of two random variables X and Y is defined by
$$\operatorname{Cov}(X,Y) := \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])].$$

Prop. 98 (Properties of Covariance)
For random variables X,Y and constants a,b:
1. $\operatorname{Cov}(X,Y)=\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]$.
2. $\operatorname{Cov}(X+a,\,Y+b)=\operatorname{Cov}(X,Y)$.
3. $\operatorname{Cov}(aX,\,bY)=ab\operatorname{Cov}(X,Y)$.
4. $\operatorname{Cov}(X,X)=\operatorname{Var}(X)$.
5. If X and Y are independent, then $\operatorname{Cov}(X,Y)=0$. (But the converse need not hold.)
6. $\operatorname{Cov}\big(\sum_i a_iX_i,\ \sum_j b_jY_j\big) = \sum_{i,j} a_ib_j\operatorname{Cov}(X_i,Y_j)$.

Thm. 99 (Law of Total Covariance)
Let X,Y be r.v.'s with finite $\mathbb{E}[X^2],\mathbb{E}[Y^2]$, and let Z be another r.v.. Then
$$\operatorname{Cov}(X,Y) = \mathbb{E}\big[\operatorname{Cov}(X,Y\mid Z)\big] + \operatorname{Cov}\big(\mathbb{E}[X\mid Z],\,\mathbb{E}[Y\mid Z]\big).$$

Def. 100 (Correlation Coefficient)
The correlation coefficient of two random variables X and Y is defined by
$$\rho(X,Y) := \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}},$$
which normalizes the covariance by the standard deviations of X and Y.

Prop. 101 (Properties of Correlation)
Let
$$\hat X := \frac{X-\mathbb{E}[X]}{\sqrt{\operatorname{Var}(X)}} \quad\text{and}\quad \hat Y := \frac{Y-\mathbb{E}[Y]}{\sqrt{\operatorname{Var}(Y)}},$$
so that $\hat X$ and $\hat Y$ have mean 0 and variance 1, then
1. $\rho(X,Y)=\operatorname{Cov}(\hat X,\hat Y)$.
2. For any $a,b>0$, $\rho(aX,bY)=\rho(X,Y)$.
3. $\rho(X,Y)\in[-1,1]$.
4. $\rho(X,Y)=1$ if $Y=aX$ for some $a>0$, and $\rho(X,Y)=-1$ if $Y=aX$ for some $a<0$.

Prop. 102 (Mean and Variance of Sums of r.v.'s)
Let $X_1,\dots,X_n$ be r.v.'s with finite $\mathbb{E}[|X_i|]$ for each i. Then
$$\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n\mathbb{E}[X_i],$$
and
$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n\operatorname{Var}(X_i) + 2\sum_{i<j}\operatorname{Cov}(X_i,X_j)$$

Thm. 103 (Sum of Independent c.r.v.'s)
Let X and Y be independent c.r.v.'s with p.d.f.'s f_X and f_Y, respectively. Then $Z:=X+Y$ has p.d.f.
$$f_{X+Y}(z) = (f_X * f_Y)(z) := \int_{\mathbb{R}} f_X(z-y)\,f_Y(y)\,dy.$$

Thm. 104 (Multi-dim. Change of Variables)
Let $\vec X=(X_1,X_2)$ have joint p.d.f. $f_{\vec X}$, and $\vec g=(g_1,g_2):\mathbb{R}^2\to\mathbb{R}^2$ be invertible with inverse $\vec h=(h_1,h_2)$. Then $\vec Y=(g_1(\vec X),g_2(\vec X))$ has
$$f_{\vec Y}(\vec y) = f_{\vec X}(\vec x)\,|J_{\vec g}(\vec x)|^{-1} = f_{\vec X}(\vec x)\,|J_{\vec h}(\vec y)|,\quad \vec x=\vec h(\vec y),$$
where
$$J_{\vec g}(\vec x) = \det\begin{bmatrix}\frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2}\\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2}\end{bmatrix} = \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1},$$
is called the *Jacobian* of $\vec g$ at $\vec x$, and $J_{\vec h}(\vec y)$ is defined similarly.

Prop. 105 (Example)
Suppose (X_1,X_2) has joint p.d.f. $f_{\vec X}(\vec x)$. Find the p.d.f. of $\vec Y=(X_1+X_2,\ X_1-X_2)$.
Step 1. Transformation: $y_1=g_1(x)=x_1+x_2$, $y_2=g_2(x)=x_1-x_2$.
Step 2. Range: all $\mathbb{R}^2$.
Step 3. Inverse:
$$x_1=h_1(y)=\tfrac12(y_1+y_2),\quad x_2=h_2(y)=\tfrac12(y_1-y_2).$$
Step 4. Jacobian:
$$J_{\vec h}(y) = \begin{vmatrix}\frac12 & \frac12\\ \frac12 & -\frac12\end{vmatrix} = -\tfrac12,\quad |J_{\vec h}(y)| = \tfrac12.$$
Step 5. Final answer:
$$f_{\vec Y}(y_1,y_2) = f_{\vec X}\left(\tfrac{y_1+y_2}{2},\ \tfrac{y_1-y_2}{2}\right)\cdot\tfrac12.$$

Thm. 106 (Change of Variables in $\mathbb{R}^n$)
Assume $\vec X=(X_1,\dots,X_n)\in\mathbb{R}^n$ has p.d.f. $f_{\vec X}(x_1,\dots,x_n)$ and $\vec Y=\vec g(\vec X)=(g_1(\vec X),\dots,g_n(\vec X))$, where $\vec g:\mathbb{R}^n\to\mathbb{R}^n$ is invertible with inverse $\vec h$. Then for $\vec y$ in the range of $\vec g$,
$$f_{\vec Y}(\vec y) = f_{\vec X}(\vec h(\vec y))\,|J_{\vec g}(\vec h(\vec y))|^{-1} = f_{\vec X}(\vec h(\vec y))\,|J_{\vec h}(\vec y)|,$$
where
$$J_{\vec g}(\vec x) = \det\begin{bmatrix}\frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_n}\end{bmatrix}(\vec x),$$
is called the *Jacobian* of $\vec g$ at $\vec x$, with $J_{\vec h}(\vec y)$ defined similarly.

Multivariate Standard Normal
$$\vec X=(X_1,\dots,X_n)^\top,\quad X_i\overset{\text{i.i.d.}}{\sim}N(0,1),$$
$$f_{\vec X}(\vec x) = \frac{1}{(2\pi)^{n/2}}\exp\left(-\tfrac12\|\vec x\|^2\right).$$

Linear Transform of Standard Normals
$$\vec Y = A\vec X,\quad A\in\mathbb{R}^{n\times n},\ \det A\ne 0,$$
$$f_{\vec Y}(\vec y) = \frac{1}{(2\pi)^{n/2}|\det A|}\exp\left(-\tfrac12\|A^{-1}\vec y\|^2\right).$$

Covariance Matrix
$$\Sigma_{\vec X} = \mathbb{E}\Big[(\vec X-\vec\mu)(\vec X-\vec\mu)^\top\Big],\quad \Sigma_{ij}=\operatorname{Cov}(X_i,X_j).$$
If X_i independent,
$$\Sigma = \operatorname{diag}(\operatorname{Var}(X_1),\dots,\operatorname{Var}(X_n)).$$

Covariance Under Linear Maps
$$\vec Y = A\vec X \ \Rightarrow\ \Sigma_{\vec Y} = A\Sigma_{\vec X}A^\top.$$
If $\vec X$ is standard normal,
$$\Sigma_{\vec Y} = AA^\top.$$

Multivariate Normal Distribution
$$\vec X\sim\mathcal{N}(\vec\mu,\Sigma) \iff$$
$$f_{\vec X}(\vec x) = \frac{1}{\sqrt{(2\pi)^n\det\Sigma}}\exp\left[-\tfrac12(\vec x-\vec\mu)^\top\Sigma^{-1}(\vec x-\vec\mu)\right].$$

Affine Transform of a Normal Vector
If
$$\vec X\sim\mathcal{N}(\vec\mu,\Sigma_{\vec X}),\qquad \vec Y = A\vec X+\vec v,$$
then
$$\vec Y\sim\mathcal{N}(A\vec\mu+\vec v,\ A\Sigma_{\vec X}A^\top).$$

Representation Using Standard Normals
Let Z_1,Z_2 be i.i.d. $N(0,1)$ and
$$\Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12} & \Sigma_{22}\end{pmatrix}.$$
Set
$$a=\frac{\Sigma_{12}}{\sqrt{\Sigma_{11}}},\qquad b=\sqrt{\Sigma_{22}-\frac{\Sigma_{12}^2}{\Sigma_{11}}}.$$
Then
$$X_1=\mu_1+\sqrt{\Sigma_{11}}\,Z_1,\qquad X_2=\mu_2+aZ_1+bZ_2,$$
and $(X_1,X_2)\sim\mathcal{N}(\mu,\Sigma)$.

Cond. Dist. of a Bivariate Normal
Let
$$(X_1,X_2)^\top\sim\mathcal{N}\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix},\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12} & \Sigma_{22}\end{pmatrix}\right).$$
Then
$$X_1\mid X_2=x\sim\mathcal{N}\left(\mu_1+\frac{\Sigma_{12}}{\Sigma_{22}}(x-\mu_2),\ \Sigma_{11}-\frac{\Sigma_{12}^2}{\Sigma_{22}}\right).$$

Quadratic Form Identity
$$\vec x^\top A\vec x+\vec v^\top\vec x = (\vec x-\vec\mu)^\top A(\vec x-\vec\mu)+\vec C \iff \vec v=-2A\vec\mu.$$

Prop. 107 (Bivariate Normal Examples)
Q1. Let $(X_1,X_2)^\top\sim N_2(\mu,\Sigma)$ with
$$\mu=\begin{pmatrix}1\\1\end{pmatrix},\qquad \Sigma=\begin{pmatrix}1 & -1\\ -1 & 4\end{pmatrix}.$$

1. **Joint distribution.**
By definition,
$$(X_1,X_2)^\top\sim N_2\left(\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}1 & -1\\ -1 & 4\end{pmatrix}\right).$$
The density is
$$\frac{1}{2\pi\sqrt{\det(\Sigma)}}\exp\left(-\frac12(\mathbf{x}-\mu)^\top\Sigma^{-1}(\mathbf{x}-\mu)\right).$$
2. **Marginals.**
For a multivariate normal, each component is univariate normal with mean and variance from μ and the diagonal of Σ:
$$\mu_1=1,\ \Sigma_{11}=1,\qquad \mu_2=1,\ \Sigma_{22}=4.$$
Hence
$$X_1\sim N(1,1),\qquad X_2\sim N(1,4).$$
3. **Conditional distribution.**
For $(X_1,X_2)^\top\sim N_2(\mu,\Sigma)$,
$$X_1\mid X_2=x\sim N\big(m(x),v\big)$$
with
$$m(x)=\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(x-\mu_2),$$
$$v=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$
Here
$$\mu_1=\mu_2=1,\quad \Sigma_{11}=1,\quad \Sigma_{22}=4,$$
$$\Sigma_{12}=\Sigma_{21}=-1.$$
Thus
$$m(x)=1+(-1)\cdot\frac14(x-1)=1-\frac{x-1}{4}=\frac{5-x}{4},$$
and
$$v=1-(-1)\cdot\frac14\cdot(-1)=1-\frac14=\frac34.$$
Therefore
$$X_1\mid X_2=x\sim N\left(\frac{5-x}{4},\ \frac34\right),\quad x\in\mathbb{R}.$$
4. **Representation using $Z_1,Z_2\sim N(0,1)$ i.i.d.**
Here
$$\mu=\begin{pmatrix}1\\1\end{pmatrix},\qquad \Sigma=\begin{pmatrix}1 & -1\\ -1 & 4\end{pmatrix},$$
so
$$\Sigma_{11}=1,\quad \Sigma_{12}=-1,\quad \Sigma_{22}=4.$$
Using
$$X_1=\mu_1+\sqrt{\Sigma_{11}}\,Z_1,\qquad X_2=\mu_2+aZ_1+bZ_2,$$
with
$$a=\frac{\Sigma_{12}}{\sqrt{\Sigma_{11}}}=-1,\qquad b=\sqrt{\Sigma_{22}-\frac{\Sigma_{12}^2}{\Sigma_{11}}}=\sqrt3,$$
we obtain
$$X_1=1+Z_1,\qquad X_2=1-Z_1+\sqrt3\,Z_2.$$
5. **$P(X_1+X_2>2)$ and $X_1\mid X_1+X_2=x$.**
Let $S=X_1+X_2$. Then S is normal with
$$\mathbb{E}[S]=\mathbb{E}[X_1]+\mathbb{E}[X_2]=1+1=2,$$
$$\operatorname{Var}(S)=\operatorname{Var}(X_1)+\operatorname{Var}(X_2)+2\operatorname{Cov}(X_1,X_2)=3.$$
Thus
$$S\sim N(2,3),\quad P(X_1+X_2>2)=P(S>2)=\frac12.$$
For the conditional distribution, consider (X_1,S). We have
$$\mathbb{E}[X_1]=1,\quad \mathbb{E}[S]=2,$$
$$\operatorname{Cov}(X_1,S)=\operatorname{Var}(X_1)+\operatorname{Cov}(X_1,X_2)=0.$$
So
$$\operatorname{Cov}\begin{pmatrix}X_1\\ S\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & 3\end{pmatrix}.$$
Since (X_1,S) is jointly normal with zero covariance, X_1 and S are independent. Thus conditioning on $S=x$ does not change the distribution of X_1:
$$X_1\mid(X_1+X_2=x)\sim N(1,1),\quad\forall x\in\mathbb{R}.$$

Q2. Suppose $\mathbf{X}=(X_1,X_2)^\top$ has p.d.f.
$$C\exp\left[-\frac12\left(2x_1^2-4x_1x_2+3x_2^2+3x_1-5x_2\right)\right].$$
We compare with the general quadratic form
$$-\frac12\left(x^\top Kx+\ell^\top x\right),\quad x=\begin{pmatrix}x_1\\x_2\end{pmatrix}.$$
- Quadratic part:
$$x^\top Kx=2x_1^2-4x_1x_2+3x_2^2.$$
For
$$K=\begin{pmatrix}k_{11} & k_{12}\\ k_{12} & k_{22}\end{pmatrix},$$
$$x^\top Kx=k_{11}x_1^2+2k_{12}x_1x_2+k_{22}x_2^2.$$
Matching coefficients:
$$k_{11}=2,\quad 2k_{12}=-4\Rightarrow k_{12}=-2,\quad k_{22}=3.$$
Thus
$$K=\Sigma^{-1}=\begin{pmatrix}2 & -2\\ -2 & 3\end{pmatrix}.$$
- Linear part:
$$\ell^\top x=3x_1-5x_2\ \Rightarrow\ \ell=\begin{pmatrix}3\\-5\end{pmatrix}.$$

For a normal $N_2(\mu,\Sigma)$,
$$-\frac12(x-\mu)^\top\Sigma^{-1}(x-\mu)$$
$$=-\frac12\left(x^\top\Sigma^{-1}x-2\mu^\top\Sigma^{-1}x+\mu^\top\Sigma^{-1}\mu\right),$$
so the linear term in x is $-2\Sigma^{-1}\mu$. Hence
$$-2\Sigma^{-1}\mu=\ell\ \Rightarrow\ \Sigma^{-1}\mu=-\frac12\begin{pmatrix}3\\-5\end{pmatrix}=\begin{pmatrix}-3/2\\5/2\end{pmatrix}.$$
First invert K:
$$K=\begin{pmatrix}2 & -2\\ -2 & 3\end{pmatrix},\quad \det(K)=2,$$
$$K^{-1}=\frac{1}{\det(K)}\begin{pmatrix}3 & 2\\ 2 & 2\end{pmatrix}=\frac12\begin{pmatrix}3 & 2\\ 2 & 2\end{pmatrix}=\begin{pmatrix}3/2 & 1\\ 1 & 1\end{pmatrix}.$$
Thus
$$\Sigma=K^{-1}=\begin{pmatrix}3/2 & 1\\ 1 & 1\end{pmatrix}.$$
Now
$$\mu=\Sigma\left(-\frac{\ell}{2}\right)=\begin{pmatrix}3/2 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}-3/2\\5/2\end{pmatrix}=\begin{pmatrix}1/4\\1\end{pmatrix}.$$
Therefore,
$$\mathbb{E}[\mathbf{X}]=\mu=\begin{pmatrix}1/4\\1\end{pmatrix},\qquad \operatorname{Cov}(\mathbf{X})=\Sigma=\begin{pmatrix}3/2 & 1\\ 1 & 1\end{pmatrix}.$$

Q3. Let $X_1,X_2\sim N(0,1)$ independent, and define
$$R=\sqrt{X_1^2+X_2^2}.$$
We want $\mathbb{E}[R]$.
The joint density of (X_1,X_2) is
$$f_{X_1,X_2}(x_1,x_2)=\frac{1}{2\pi}\exp\left(-\frac{x_1^2+x_2^2}{2}\right).$$
Change to polar coordinates:
$$f_{R,\Theta}(r,\theta)=\frac{1}{2\pi}\exp\left(-\frac{r^2}{2}\right)\cdot r.$$
Integrating out θ gives
$$f_R(r)=\int_0^{2\pi}f_{R,\Theta}(r,\theta)\,d\theta=re^{-r^2/2},\quad r\ge0.$$
$$\mathbb{E}[R]=\int_0^\infty r\cdot f_R(r)\,dr=\int_0^\infty r^2e^{-r^2/2}\,dr.$$
Use substitution $u=\frac{r^2}{2}$, so $r^2=2u$, $dr=\frac{1}{\sqrt{2u}}\,du$:
$$\int_0^\infty r^2e^{-r^2/2}\,dr=\int_0^\infty(2u)e^{-u}\frac{1}{\sqrt{2u}}\,du=\sqrt2\int_0^\infty u^{1/2}e^{-u}\,du.$$
Recognize the Gamma function:
$$\int_0^\infty u^{1/2}e^{-u}\,du=\Gamma\left(\frac32\right)=\frac12\sqrt\pi.$$
Therefore
$$\mathbb{E}[R]=\sqrt2\cdot\Gamma\left(\frac32\right)=\sqrt2\cdot\frac12\sqrt\pi=\sqrt{\frac\pi2}.$$
So
$$\boxed{\mathbb{E}\big[\sqrt{X_1^2+X_2^2}\big]=\sqrt{\frac\pi2}.}$$

Thm. 108 (Covariance of Linear Combinations)
For real constants a_i,b_j and random variables X_i,Y_j,
$$\operatorname{Cov}\left(\sum_{i=1}^m a_iX_i,\ \sum_{j=1}^n b_jY_j\right)=\vec a^\top\,\mathbb{E}[\vec X\vec Y^\top]\,\vec b.$$

Def. 109 (Probability Generating Function)
Let X be a non-negative integer-valued r.v. with $\Pr(X=k)=p_k$, $k=0,1,2,\dots$. The *probability generating function* (PGF) of X is
$$G_X(s):=\mathbb{E}[s^X]=\sum_{k=0}^\infty p_ks^k,\qquad |s|\le1.$$
G_X is analytic for $|s|<1$ and encodes the entire p.m.f. $\{p_k\}$.

Thm. 110 (Recovery of Moments)
If $\mathbb{E}[X]<\infty$, then
$$\mathbb{E}[X]=G_X'(1),\qquad \mathbb{E}[X(X-1)]=G_X''(1).$$
Hence $\operatorname{var}(X)=G_X''(1)+G_X'(1)-[G_X'(1)]^2$.

Prop. 111 (Additivity for Independent Sums)
If X and Y are independent integer-valued r.v.'s with PGFs G_X and G_Y, then
$$G_{X+Y}(s)=G_X(s)\,G_Y(s),\qquad |s|\le1.$$

Prop. 112 (Tail Generating Function)
Let $t(n)=\Pr(X>n)$ and $T(s)=\sum_{n\ge0}t(n)s^n$. Then
$$T(s)=\frac{1-G_X(s)}{1-s},\qquad \mathbb{E}[X]=T(1),$$
$$\operatorname{Var}(X)=2T'(1)+T(1)-T(1)^2.$$

Prop. 113 (Common PGFs)
$$X\sim\operatorname{Ber}(p)\ \Rightarrow\ G_X(s)=(1-p)+ps,$$
$$X\sim\operatorname{Bin}(n,p)\ \Rightarrow\ G_X(s)=(1-p+ps)^n,$$
$$X\sim\operatorname{Geom}(p)\ \Rightarrow\ G_X(s)=\frac{p}{1-(1-p)s},$$
$$X\sim\operatorname{Pois}(\lambda)\ \Rightarrow\ G_X(s)=\exp[\lambda(s-1)].$$

Def. 114 (Moment Generating Function)
The *moment generating function* (MGF) of a r.v. X is
$$M_X(t):=\mathbb{E}[e^{tX}],\qquad t\in\mathbb{R}.$$
M_X is well-defined if $M_X(t)<\infty$ for all $t\in(-\eta,\eta)$ for some $\eta>0$. Note that $0\in\{t\in\mathbb{R}:M_X(t)<\infty\}$.

Thm. 115
If for some $\delta>0$, $\mathbb{E}[e^{tX}]<\infty$ for all $|t|<\delta$, then
$$\mathbb{E}[X^k]=M_X^{(k)}(0)=\left.\frac{d^kM_X(t)}{dt^k}\right|_{t=0}.$$

Prop. 116 (Multiplicative Property)
If X and Y are independent with well-defined MGFs $M_X(t)$ and $M_Y(t)$ for $|t|<\delta$, then
$$M_{X+Y}(t)=M_X(t)\,M_Y(t),\qquad t\in(-\delta,\delta).$$

Prop. 117 (Uniqueness Property)
Suppose X and Y have MGFs M_X and M_Y, respectively. If there exists $\delta>0$ such that
$$M_X(t)=M_Y(t),\quad\forall t\in(-\delta,\delta),$$
then X and Y have the same distribution (i.e. the same c.d.f. and p.d.f.).

Prop. 118 (Common MGFs)
$$X\sim\operatorname{Ber}(p)\ \Rightarrow\ M_X(t)=\mathbb{E}[e^{tX}]=1-p+pe^t,$$
$$X\sim\operatorname{Bin}(n,p)\ \Rightarrow\ M_X(t)=(1-p+pe^t)^n,$$
$$X\sim\operatorname{Geom}(p)\ \Rightarrow\ M_X(t)=\frac{pe^t}{1-(1-p)e^t},$$
$$X\sim\operatorname{Exp}(\lambda)\ \Rightarrow\ M_X(t)=\frac{\lambda}{\lambda-t},$$
$$X\sim\operatorname{Pois}(\lambda)\ \Rightarrow\ M_X(t)=\exp[\lambda(e^t-1)],$$
$$X\sim\mathcal{N}(\mu,\sigma^2)\ \Rightarrow\ M_X(t)=\exp\left(\mu t+\tfrac12\sigma^2t^2\right).$$