Contents

Probabiliy and Statistics

Partition

A *Partition* of Ω is a collection of nonempty subsets $A_1, ..., A_n$ in Ω such that

1. the A_j are *exhaustive* i.e., $A_1 \cup A_2 \cup ... \cup A_n = \Omega$
2. the A_j are *disjoint* i.e., $A_i \cap A_j = \emptyset$, for $i \neq j$

Combinatorics

- Given n objects, the nummber of different **permutations (without repetition)** of length $ r   n $ is

$$n(n-1)(n-2)...(n-r+1) = \frac{n!}{(n-r)!}$$

- Given $n = \sum_{i=1}^{r} n_i$ objects of r different types, where n_i is the number of objects of type i that are **indistinguishable** from one another, the number of **permutations (without repetition)** of the n objects is

$$\frac{n!}{n_1!n_2!...n_r!}$$

- **Binomial coefficient**
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$
- the number of ways of choosing a set of r objects from a set of n distinct objects without repetition (and order does not matter) is
$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$
this is essentialy the same as the permutation (without repetition) except that one divides by the number of permutations of the r objects of r since order does not matter.
- The number of ways of distributing n distinct objects into r distinct groups of size $n_1, ..., n_r$, where $n_1 + ... + n_r = n$, is
$$\frac{n!}{n_1!n_2!...n_r!}$$
one actually does not care about the order in the groups, that's why one divides by the number of permutations in each group.
- The number of distinct vectors $(n_1, ..., n_r)$ of positive integers, $n_1, ..., n_r > 0$, satisfying $n_1 + ... + n_r = n$, is
$$\binom{n-1}{r-1}$$
- The number of distinct vectors $(n_1, ..., n_r)$ of positive integers, $n_1, ..., n_r 0$, satisfying $n_1 + ... + n_r = n$, is
$$\binom{n+r-1}{n} = \binom{n+r-1}{r-1}$$
example: How many different ways are there to put 6 identical balls in 3 boxes? First align $3-1$ delimiters and the 6 balls:
$$||oooooo$$
Now the number of permutations (if we consider the objects different from each other) is 8!. However, since there are only two types of indistinguishable objects (bars, and balls), as seen above, one has to divide by the number of permutations of the 6 balls (6!) and the number of permutations of the 2 bars (2!). so:
$$\frac{8!}{6!2!} = \binom{6+3-1}{6} = \binom{6+3-1}{3-1} = 28$$

Permutation: you care about the *order* of *selection*

Combination: you don't care about the order of *selection*

Probabilities

Probability space $(\Omega, \mathcal{F}, P)$:

1. Ω: *sample space* (contains all possible results)
2. $\mathcal{F}$: *event space* (represents the set of events where each event is a *set* of outcomes; it is the *power set* of the sample space and it's cardinal is $2^{|\Omega|}$)
3. P: *probability distribution* (which associates a probability $P(A) \in [0, 1]$ to each $A \in \mathcal{F}$)
   if $\{A_i\}_{i=1}^{\infty}$ are pairwise *disjoint* (i.e., $A_i \cap A_j = \emptyset, i \neq j$), then
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P\big(\bigcup_{i=1}^{\infty} A_i \sum_{i=1}^{\infty} P(A_i)\big)$ (Boole's inequality)

Inclusion-exclusion formulae

- $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$
- first add all the "unique" probabilities, then substract all the possible pairs, add all the possible triples, substract all the possible quadruplets, etc. (always alternate signs).
- the number of terms in the general formula is $2^n - 1$

Conditional probability

- $P(A|B) = \frac{P(A \cap B)}{P(B)}$ (since B is given, the sample space shrinks to B), you can extend this definition on more than two events by *recursion*.

Law of total probability

$$P(A) = \sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i)$$

where $\{B_i\}_{i=1}^{\infty}$ are pairwise disjoint events, and $A \subset \bigcup_{i=1}^{\infty} B_i$.

Bayes

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{\infty} P(A|B_i)P(B_i)}$$

assuming the same conditions as above.

Independence

$$P(A|B) = P(A)$$

Two events are independent *iff*

$$P(A \cap B) = P(A)P(B)$$

One can have mutually independent events, pairwise independent events and conditionally independent events (see course book for more details)

Random variables

- Let (Ω, F, P) be a probability space. A *random variable (rv)* $\Omega \to \mathbb{R}$ is a function from the sample space Ω taking values in the real numbers $\mathbb{R}$ (A mapping from the sample space to the reals).
- The set of values taken by X,

  $$D_X = \{x \in \mathbb{R} : \exists \omega \in \Omega \textit{ such that } X(\omega) = x\}$$

  is called the support of X. If D_X is countable, then X is a *discrete random variable*.
- In particular we set $A_x = \{\omega \in \Omega : X(\omega) = x\}$
- A random variable that takes only the values **0** and **1** is called an **indicator variable**, or a **Bernoulli random variable**, or a Bernoulli trial

Mass functions (PMF)

The **Probability Mass Function (PMF)** of a discrete random variable X is

$$f_X(x) = P(X = x) = P(A_x), x \in \mathbb{R}.$$

It has two key properties:

1. $f_X(x)0$, and it is only positive for $x \in D_X$, where D_X is the image of the function X, i.e., the *support* of f_X;
2. The total probability is equal to 1.

Distributions

Binomial random variable A *binomial* random variable X has PMF

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 0, 1, ..., n,\ n \in \mathbb{N},\ 0p1$$

With $n = 1$, this is a Bernoulli variable.

- The binomial model is used when we are considering the number of 'successes' of a trial which is independently repeated a fixed number of times, and where each trial has the same probability of success.
- $\mathbb{E}(X) = np$
- $Var(X) = np(1-p)$

Geometric random variable A *geometric* random variable has PMF

$$f_X(x) = p(1-p)^{x-1}\ x = 1, 2...\ 0p1.$$

- This models the waiting time until a first event, in a series of independent trials having the same success probability.
- $\mathbb{E}(X) = \frac{1}{p}$
- $Var(X) = \frac{1-p}{p^2}$

Negative binomial random variable A *negative binomial* random variable X has PMF

$$f(x) = \binom{x-1}{n-1} p^n (1-p)^{x-n},\ x = 0, 1, ..., n,\ n \in \mathbb{N},\ 0p1$$

With $n = 1$, this is a Geometric random variable.

- It models the waiting time until the nth success in a series of independent trials having the same success probability.
- $\mathbb{E}(X) = \frac{n}{p}$
- $Var(X) = \frac{n(1-p)}{p^2}$

(Gamma function)

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1}e^{-u}du,\ \alpha > 0$$

$$\Gamma(1) = 1$$

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha),\ \alpha > 0$$

$$\Gamma(n) = (n-1)!, n = 1, 2, 3, ...$$

$$\Gamma(1/2) = \sqrt{\pi}$$

Hypergeometric random variable We draw a sample of m balls *without replacement* from an urn containing w white balls and b black balls. Let X be the number of white balls drawn. Then

$$P(X = x) = \frac{\binom{w}{x}\binom{b}{m-x}}{\binom{w+b}{m}} \quad x = max(0, m-b), ..., min(w, m)$$

Discrete uniform random variable A discrete uniform random variable X has PMF

$$f_X(x) = \frac{1}{b-a+1},\ x = a, a+1, ..., b,\ a < b,\ a, b \in \mathbb{Z}$$

Poisson random variable A Poisson random variable X has PMF

$$f_X(x) = \frac{\lambda^x}{x!}e^{-\lambda},\ x = 0, 1, ...,\ \lambda > 0.$$

- $\mathbb{E}(X) = \lambda$
- $Var(X) = \lambda$

For a function to be a probability distribution

$$\sum_{x=0}^{\infty} f_X(x) == 1$$

Cumulative distribution function (CDF)

The **Cumulative Distribution function (CDF)** of a random variable X is

$$F_X(x) = P(Xx),\ x \in \mathbb{R}$$

If X is discrete, we can write

$$F_X(x) = \sum_{\{x_i \in D_X : x_i x\}} P(X = x_i),$$

which is a step function with jumps at the points of the support D_X of $f_X(x)$

Transformation of discrete random variables

$$f_Y(y) = P(Y = y) = \sum_{x:g(x)=y} P(X = x) = \sum_{x:g(x)=y} f_X(x)$$

$$Y = g(X)$$

Expectation

$$E(X) = \sum_{x \in D_X} xP(X = x) = \sum_{x \in D_X} xf_X(x)$$

analogous to the center of mass in physics, is also sometimes called "*average of X*"

Expected value of a function Let X be a random variable with mass function f, and let g be a real-valued function of $\mathbb{R}$. Then

$$E\{g(X)\} = \sum_{x \in D_X} g(x)f(x)$$

If $g(x) = x$ we have the normal expectation.

Properties of the expected value

1. $E(aX + b) = aE(X) + b(linearity)$
2. $E\{g(X) + h(X)\} = E\{g(X)\} + E\{f(X)\}$
3. if $P(X = b) = 1$, then $E(X) = b$
4. if $P(a < Xb) = 1$, then $a < E(X)b$
5. $\{E(X)\}^2 E(X^2)$

Variance

$$var(X) = E\big[\{X - E(X)\}^2\big]$$

Average squared distance of X (moment of inertia)

Standard deviation

$$stdDev(X) = \sqrt{var(X)}$$

Properties of variance $var(X) = E(X^2) - E(X)^2 = E\{X(X-1)\} + E(X) - E(X)^2$

$var(aX + b) = a^2 var(X)$

$var(X) = 0 \implies X$ is constant with probability 1.

$$E(X) = \sum_{x=1}^{\infty} P(Xx)$$

Conditional probability distributions

Let (Ω, F, P) be a probability space, on which we define a random variable X, and let $B \in F$ with $P(B) > 0$. Then the **conditional probability mass function** of X given B is

$$f_X(x|B) = P(X = x|B) = P(A_x \cap B)/P(B)$$

where $A_x = \{\omega \in \Omega : X(\omega) = x\}$

Conditional expected value The conditional expected value of $g(X)$ given B is

$$E\{g(X)|B\} = \sum_x g(x) f_X(x|B)$$

Let X be a random variable with expected value $E(X)$ and let B be an event with $P(B), P(B^c) > 0$. Then

$$E(X) = E(X|B)P(B) + E(X|B^c)P(B^c)$$

More generally, when $\{B_i\}_{i=1}^{\infty}$ is a partition of Ω, $P(B_i) > 0$ for all i, and the sum is absolutely convergent,

$$E(X) = \sum_{i=1}^{\infty} E(X|B_i)P(B_i)$$

Notions of Convergence Let $\{X_n\}$, X be random variables whose cumulative distribution functions are $\{F_n\}$, F. Then we say that the random variables $\{X_n\}$ **converge in distribution** to X, if, for all $x \in \mathbb{R}$ where F is continuous,

$$F_n(x) \to F(x), \ n \to \infty$$

We write $X_n \xrightarrow{D} X$

Lemma: $n^{-r}\binom{n}{r} \to 1/r!$ for all $r \in \mathbb{N}$, when $n \to \infty$

Law of small numbers:

Let $X_n \sim B(n, p_n)$, and suppose that $np_n \to \lambda > 0$ when $n \to \infty$. Then $X_n \to X$, where $X \sim Pois(\lambda)$

Continuous Random Variables

A random variable X is *continuous* if there exists a function $f(x)$, called **probability density function (PDF)** (If X is discrete, then its PMF $f(x)$ is often also called its density function) of X, such that

$$P(Xx) = F(x) = \int_{-\infty}^{x} f(u)du, \ x \in \mathbb{R}$$

Uniform distribution

$$f(u) = \begin{cases} \frac{1}{b-a}, & aub \\ 0 & \text{otherwise} \end{cases} \quad a < b$$

We write $U \sim U(a, b)$

$$F(u) = \begin{cases} 0 & u < a \\ \frac{u-a}{b-a} & aub \\ 1 & x > b \end{cases}$$

Exponential distribution

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0 & \text{otherwise} \end{cases}$$

We write $X \sim exp(\lambda)$

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x0 \\ 0 & x < 0 \end{cases}$$

$\mathbb{E}(X) = \frac{1}{\lambda}$
$Var(X) = \frac{1}{\lambda^2}$

Gamma distribution

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0, \\ 0 & \text{otherwise} \end{cases}$$

We write $X \sim Gamma(a, \lambda)$

α is the *shape parameter*

λ is the *rate*

λ^{-1} is the *scale parameter*

Laplace distribution

$$f(x) = \frac{\lambda}{2} e^{-\lambda |x-\eta|}, \ x \in \mathbb{R}, \ \eta \in \mathbb{R}, \ \lambda > 0$$

Pareto distribution

$$F(x) = \begin{cases} 0, & x < \beta, \\ 1 - \left(\frac{\beta}{x}\right)^\alpha & x\beta, \end{cases}, \alpha, \beta > 0$$

$$f(x) = \begin{cases} 0, & x < \beta, \\ \frac{\alpha \beta^\alpha}{x^{\alpha+1}} & x\beta, \end{cases}, \alpha, \beta > 0$$

Moments

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx,$$

$$var(X) = \int_{-\infty}^{\infty} \{x - E(X)\}^2 f(x) dx = E(X^2) - E(X)^2$$

Conditional densities (quite useful)**

$$F_X(x|X \in A) = P(Xx|X \in A) = \frac{P(Xx \ \cap \ X \in A)}{P(X \in A)} = \frac{\int_{A_x} f(y) dy}{P(X \in A)}$$

where $A_x = \{y : yx, \ y \in A\}$

$$f_X(x|X \in A) = \begin{cases} \frac{f_X(x)}{P(X \in A)}, & x \in A, \\ 0, & \text{otherwise} \end{cases}$$

Quantile Let $0 < p < 1$. We define the p quantile of the cumulative distribution function $F(x)$ to be

$$x_p = inf\{x : F(x)p\}$$

For most continuous random variables, x_p is unique and equals $x_p = F^{-1}(p)$, where F^{-1} is the inverse function F; the x_p is the value for which $P(Xx_p) = p$.

General transformation Let $g : \mathbb{R} \to \mathbb{R}$ be a function and $\mathcal{B} \subset \mathbb{R}$ any subset of $\mathbb{R}$. Then $g^{-1}(\mathcal{B}) \subset \mathbb{R}$ is the set for which $g\{g^{-1}(\mathcal{B})\} = \mathcal{B}$.

Let $Y = g(X)$ be a random variable and $\mathcal{B}_y = (-\infty, y]$. Then

$$F_Y(y) = P(Y \leq y) = \begin{cases} \int_{g^{-1}(\mathcal{B}_y)} f_X(x)dx, & \text{X continuous,} \\ \sum_{x \in g^{-1}(\mathcal{B}_y)} f_X(x), & \text{X discrete,} \end{cases}$$

where $g^{-1}(\mathcal{B}_y) = \{x \in \mathbb{R} : g(x)y\}$. When g is monotone increasing or decreasing and has inverse g^{-1}, then

$$f_Y(y) = \left| \frac{dg^{-1}(y)}{dy} \right| f_X\{g^{-1}(y)\},\ y \in \mathbb{R}.$$

Normal Distribution or Gaussian Distribution A random variable X having density

$$f(x) = \frac{1}{(2\pi)^{1/2}\sigma} exp\big\{-\frac{(x-\mu)^2}{2\sigma^2}\big\},\ x \in \mathbb{R},\ \mu \in \mathbb{R}, \sigma > 0$$

is a **normal random variable** with expectation μ and variance σ^2: we write $X \sim \mathcal{N}(\mu, \sigma^2)$.

When $\mu = 0,\ \sigma^2 = 1$, the corresponding random variable Z is **standard normal**, $Z \sim \mathcal{N}(0, 1)$, with density

$$\phi(z) = (2\pi)^{-1/2} e^{-z^2/2},\ z \in \mathbb{R}$$

$$F_Z(x) = P(Zx) = \Phi(x) = \int_{-\infty}^{x} \phi(z)dz = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{x} e^{-z^2/2} dz$$

The density $\phi(z)$, the cumulative distribution function $\Phi(z)$, and the quantiles z_p of $Z \sim \mathcal{N}(0, 1)$ satisfy, for all $z \in \mathbb{R}$:

Properties

1. the density is symmetric with respect to $z = 0$, i.e., $\phi(z) = \phi(-z)$;
2. $P(Zz) = \Phi(z) = 1 - \Phi(-z) = 1 - P(Zz)$;
3. the standard normal quantiles z_p satisfy $z_p = -z_{1-p}$, for all $0 < p < 1$;
4. $z^r\phi(z) \to 0$ when $z \to \pm\infty$, for all $r > 0$. This implies that the moments $\mathbb{E}(Z^r)$ exist for all $r \in \mathbb{N}$;
5. we have
$$\phi'(z) = -z\phi(z),\ \phi''(z) = (z^2 - 1)\phi(z),\ \phi'''(z) = -(z^3 - 3z)\phi(z)$$
This implies that $\mathbb{E}(Z) = 0$, $var(Z) = 1$, $\mathbb{E}(Z^3) = 0$, etc
6. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$ (**useful**)

Normal approximation to the binomial distribution Let $X_n \sim B(n,p)$, where $0 < p < 1$, let

$$\mu = \mathbb{E}(X_n) = np,\ \sigma_n^2 = var(X_n) = np(1-p)$$

and let $Z \sim \mathcal{N}(0,1)$. When $n \to \infty$,

$$P\left(\frac{X_n - \mu_n}{\sigma_n} z\right) \to \Phi(z),\ z \in \mathbb{R}$$

This gives us an approximation of the probability that $X_n r$:

$$P(X_n r) = P\left(\frac{X_n - \mu_n}{\sigma_n} \frac{r - \mu_n}{\sigma_n}\right) \doteq \Phi\left(\frac{r - \mu_n}{\sigma_n}\right)$$

which corresponds to $X_n \dot{\sim} \mathcal{N}\{np, np(1-p)\}$. The normal approximation is valid for large n and $min\{np, n(1-p)\}5$.

A better approximation to $P(X_n r)$ is given by replacing r by $r + 1/2$; the $1/2$ is called the **continuity correction**. This gives

$$P(X_n r) \doteq \Phi\left(\frac{r + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Q-Q Plots If straightt line of 45° angle: GOOD distribution.

Several Random Variables

The (joint) cumulative distribution function of (X, Y) is

$$F_{X,Y}(x,y) = P(Xx, Yy) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)dudv,\ (x,y) \in \mathbb{R}^2$$

and this implies that

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

Marginal probability mass/density function The marginal probability mass/density function of X is

$$f_X(x) = \begin{cases} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy, & \text{continuous case,} \\ \sum_y f_{X,Y}(x,y), & \text{discrete case} \end{cases}$$

Conditional probability mass/density function The conditional probability mass/density function of Y given X is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)},\ y \in \mathbb{R}$$

Analogous for cumulative distribution. We can extend all these definitions to n random variables.

Multinomial distribution The random variable $(X_1, ..., X_k)$ has the multinomial distribution of denominator m and probabilities $(p_1, ..., p_k)$ if its mass function is

$$f(x_1, ..., x_k) = \frac{m!}{x_1!x_2!...x_k!} p_1^{x_1} p_2^{x_2} ... p_k^{x_k},\ x_1, ..., x_k \in \{0, ..., m\}, \sum_{j=1}^{k} x_j = m$$

Multivariable independence Random variables X,Y defined on the same probability space are independent if

$$P(X \in \mathcal{A}, Y \in \mathcal{B}) = P(X \in \mathcal{A})P(Y \in \mathcal{B})$$

$$F_{X,Y}(x, y) = ... = F_X(x)F_Y(y)$$

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

If X, Y are independent, then for all x such that $f_X(x) > 0$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{f_X(x)f_Y(y)}{f_X(x)}$$

Joint moments Let X, Y be random variables of density $f_{X,Y}(x, y)$. Then if $E\{|g(X, Y)|\} < \infty$, we can define the expectation of $g(X, Y)$ to be

$$E\{g(X, Y)\} = \begin{cases} \int\int g(x, y)f_{X,Y}(x, y)dxdy, & \text{continuous case,} \\ \sum_{x,y} g(x, y)f_{X,Y}(x, y), & \text{discrete case} \end{cases}$$

$$cov(X, Y) = E[\{X - E(X)\}\{Y - E(Y)\}] = E(XY) - E(X)E(Y)$$

Properties of covariance

$$cov(X, X) = var(X)$$

$$cov(a, X) = 0$$

$$cov(X, Y) = cov(Y, X)$$

$$cov(a + bX + cY, Z) = b\ cov(X, Z) + c\ cov(Y, Z)$$

$$cov(a + bX, c + dY) = bd\ cov(X, Y)$$

$$var(a + bX + cY) = b^2 var(X) + 2bc\ cov(X, Y) + c^2 var(Y)$$

$$cov(X, Y)^2 var(X)var(Y)$$

If X and Y are independent and $g(X)$, $h(Y)$ are functions whose expectations exist, then

$$E\{g(X)h(Y)\} = ... = E\{g(X)\}E\{h(Y)\}$$

$$X, Y \text{ indep} \implies cov(X, Y) = 0$$

Correlation

$$corr(X, Y) = \frac{cov(X, Y)}{\{var(X)var(Y)\}^{1/2}}$$

Measures the linear dependence between X and Y