

Probability The chance something happens We use models to represent a system Types of models: <ul style="list-style-type: none">Deterministic: Outcome is determined by input conditionsProbabilistic/Stochastic: Outcome can vary; degree of randomness Building a model: <ol style="list-style-type: none">Define the random experiment inherent in the problem,Specify the set of all possible outcomes and events of interest,Specify a probability assignment for the outcomes and events.	Family of discrete random variables Bernoulli(p) RV: <ul style="list-style-type: none">$p_X(x) = \begin{cases} 1-p, & x=0 \\ p, & x=1 \\ 0, & \text{otherwise} \end{cases}$Used for experiments with 2 outcomes Binomial(n, p) RV: <ul style="list-style-type: none">$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x=0,1,2,\dots,n \\ 0, & \text{otherwise} \end{cases}$Used in exp. with n trials of 2 outcomes Poisson(α) RV: <ul style="list-style-type: none">$p_X(x) = \begin{cases} \frac{\alpha^x e^{-\alpha}}{x!}, & x=0,1,2,\dots \\ 0, & \text{otherwise} \end{cases}, \alpha = \lambda T$Used for number of completely random events		Geometric(p) RV: <ul style="list-style-type: none">$p_X(x) = \begin{cases} p(1-p)^{x-1}, & x=1,2,3,\dots \\ 0 & \text{otherwise} \end{cases}$Used for no. of tries until desired outcome Pascal(k, p) RV: <ul style="list-style-type: none">$p_X(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}, x=k, k+1, \dots$Used for probability of k successes
Set theory In a set: $x \in A$ Not in a set: $x \notin A$ Set of tabular method: $S = \{x_1, x_2, x_3, \dots\}$ Set of rule method: $S = \{x x \text{ satisfies } P\}$ Countable and uncountable are self-explanatory Empty/null set: $\emptyset = \{\}$ Subsets: $A \subseteq B, A = \{x_1, x_2\}, B = \{x_1, x_2, x_3\}, \emptyset \subseteq A$ Universal set S with N elements, 2^N subsets of S Union: $A \cup B = \{x x \in A \text{ or } x \in B\}$ Intersection: $A \cap B = \{x x \in A \text{ and } x \in B\}$ Complement: $A^c = \{x x \notin A\}, S^c = \emptyset$ Set difference: $A - B = \{x x \in A \text{ and } x \notin B\}$ Commutative, distributive, and associative apply to union and intersections. De Morgan's law: $(A \cup B)^c = A^c \cap B^c, (A \cap B)^c = A^c \cup B^c$ Mutually exclusive (disjoint) sets: $A_i \cap A_j = \emptyset, i \neq j$ Partition: $\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \dots \cup A_n = S$	Cumulative distribution function CDF: $F_X(x) = P[X \leq x]$ CDF is cumulative, while PMF is per value CDF is 1 at the end, while PMF is not Theorem: $F_X(b) - F_X(a) = P[a < X \leq b]$	Expected Value Expected value: $E[X] = \mu_X = \sum x p_X(x)$ Bernoulli: $E[X] = p$ Binomial: $E[X] = np$ Geometric: $E[X] = \frac{1}{p}$ Poisson: $E[X] = \alpha$	
	Function of RV Function: $Y = g(X)$ For discrete, PMF is: $p_Y(y) = \sum \{x g(x) = y\} p_X(x)$ Expected value of Y : $\mu_Y = E[Y] = E[g(x)] = \sum_{x \in S_X} g(x) p_X(x)$ Linear func of RV: $E[aX + b] = aE[X] + b$ Variance: $VAR[X] = \sigma_X^2 \triangleq E[(X - \mu_X)^2] = \sum_{x \in S_X} (x - \mu_X)^2 p_X(x)$ Also variance: $VAR[X] = E[X^2] - (E[X])^2$ Also variance: $VAR[aX + b] = a^2 VAR[X] \in S_X$ Conditional PMF: $p_{X B}(x) = P[X = x B]$	Variances for discrete RVs Bernoulli: $VAR[X] = p(1-p)$ Binomial: $VAR[X] = np(1-p)$ Geometric: $VAR[X] = \frac{1-p}{p^2}$ Poisson: $VAR[X] = \alpha$	
Set to probability <div>Element - Outcome Set - Event Universal Set - Sample space</div> Probability Law: $A \subset S \Rightarrow \{P[\cdot]\} P[A] \in [0,1] \in \mathbb{R}$ Axioms of Probability: <ol style="list-style-type: none">For any event $A, P[A] \geq 0$$P[S] = 1$If $A \cap B = \emptyset$, then $P[A \cup B] = P[A] + P[B]$ Consequences of these axioms: <ol style="list-style-type: none">$P[\emptyset] = 0$$P[A^c] = 1 - P[A]$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$If $A \subseteq B$, then $P[A] \leq P[B]$ Union bound: $P[A \cup B] \leq P[A] + P[B]$	Continuous random variables For things that are not discrete CDF of continuous RV: $F_X(x) = P[X \leq x]$ Probability density function (PDF): $f_X(x) = \frac{dF_X(x)}{dx}$ PDF theorem: $P[x_1 < X \leq x_2] = \int_{x_1}^{x_2} f_X(x) dx$ Expected value: $E[X] = \mu_X = \int_{-\infty}^{\infty} x f_X(x) dx, E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$ Variance: $\sigma_X^2 = VAR[X] = E[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$		
	Family of continuous random variables Uniform(a, b) RV: $f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}, F_X(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}, E[X] = \frac{a+b}{2}, VAR[X] = \frac{(b-a)^2}{12}$ Exponential(λ) RV: $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}, F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}, E[X] = \frac{1}{\lambda}, VAR[X] = \frac{1}{\lambda^2}$ Erlang(n, λ): $f_X(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{n-1}}{\Gamma(n)} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}, \Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx (\alpha > 0), E[X] = \frac{n}{\lambda}, VAR[X] = \frac{n}{\lambda^2}$ Gaussian(μ, σ^2): $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} - \infty < x < \infty, E[X] = \mu, VAR[X] = \sigma^2$ CDF of standard normal RV Z : $F_Z(z) = P[Z \leq z] = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \triangleq \Phi(z), P[X \leq x] = F_X(x) = \Phi(\frac{x-\mu}{\sigma})$		
More probability Conditional probability: $P[A B] \triangleq \frac{P[A \cap B]}{P[B]}, P[A \cap B] = P[A B]P[B]$ a) $P[A B] \geq 0$ b) $P[S B] = 1$ c) If $A = A_1 \cup A_2 \cup \dots$ with $A_i \cap A_j = \emptyset$ for $i \neq j$, then $P[A B] = P[A_1 B] + P[A_2 B] + \dots$ Law of total probability: $P[B] = \sum_{i=1}^n P[B A_i]P[A_i]$ Bayes' Theorem: $P[A_i B] = \frac{P[B A_i]P[A_i]}{P[B]}$ Independence: $P[A \cap B] = P[A]P[B]$ Disjoint: $P[A \cap B] = P[\emptyset] = 0$ INDEPENDENT EVENTS AIN'T DISJOINT AND VICE VERSA Bernoulli trials: Find probability of an outcome after n trials Reliability Problem (series): If one fails, whole thing fails Reliability Problem (parallel): If all fails, whole thing fails	Conditioning continuous RV Cond. PDF: $f_{X B}(x) = \begin{cases} \frac{f_X(x)}{P[B]} & x \in B \\ 0 & \text{otherwise} \end{cases}, E[X B] = \int_{-\infty}^{\infty} x f_{X B}(x) dx, VAR[X B] = E[X^2 B] - E[X B]^2$ Memoryless property: Let X be an exponential RV $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}, F_X(x) = 1 - e^{-\lambda x} \quad x \geq 0$ Then we have $P[X > x + h X > x] = P[X > h]$		
	Dirac delta function & unit step Delta function: $\delta(x) = \lim_{\epsilon \rightarrow 0} d_{\epsilon}(x) = \frac{du(x)}{dx}$ Area of delta: $\int_{-\infty}^{\infty} \delta(x - x_0) dx = 1$ Unit step: $u(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} = \int_{-\infty}^x \delta(t) dt$ PMF and PDF: $p_X(x) = f_X(x) = \sum_{x_i \in S_X} p_X(x_i) \delta(x - x_i)$		
Discrete random variables Range of an RV: $S_X \subseteq \mathbb{R}$ Discrete RV: $S_X = \{x_1, x_2, \dots\}$ Continuous RV: S_X is an uncountable set Mixed RV: If it has elements of both	Probability mass function: $p_X(x) = P[X = x]$ <ol style="list-style-type: none">For any $x, p_X(x) \geq 0$$\sum_{x \in S_X} p_X(x) = 1$For any event $B \subseteq S_X, P[B] = \sum_{x \in B} p_X(x)$		
	Mixed RV: when CDF is continuous except having jumps at a number of x_i 's. Mixed PDF: $f_X(x) = f_X^c(x) + p_X(x_i) \delta(x - x_i)$ Mixed CDF: $F_X(x) = F_X^c(x) + p_X(x_i) u(x - x_i)$		

Derived RV probability models

Given X and $f_X(x)$, find the PDF of $Y = g(X)$

1. Find CDF of Y : $P[Y \leq y]$ by plugging $Y = g(X)$

2. Obtain PDF of Y as $f_Y(y) = \frac{dF_Y(y)}{dy}$

Constant addition transformation: If $Y = X + b$, we have

$F_Y(y) = F_X(y - b)$ and $f_Y(y) = f_X(y - b)$

Constant multiplication transformation: If $Y = aX$, we have

For $a > 0$: $F_Y(y) = F_X\left(\frac{y}{a}\right)$, $f_Y(y) = \frac{1}{a}f_X\left(\frac{y}{a}\right)$

For $a < 0$: $F_Y(y) = 1 - F_X\left(\frac{y}{a}\right)$, $f_Y(y) = -\frac{1}{a}f_X\left(\frac{y}{a}\right)$

General linear transformation: If $Y = aX + b$, we have

$f_Y(y) = \frac{1}{|a|}f_X\left(\frac{y - b}{a}\right)$

Quadratic transformation: If $Y = aX^2 + b$, where $a > 0$,

$f_Y(y) = f_X\left(\sqrt{\frac{y - b}{a}}\right) \times \frac{1}{2a\sqrt{\frac{y - b}{a}}} + f_X\left(-\sqrt{\frac{y - b}{a}}\right) \times \frac{1}{2a\sqrt{\frac{y - b}{a}}}$

How to generate RV Y of desired distribution $F_Y(y)$:

- 1. Generate a RV $X \sim \text{Uniform}(0, 1)$
- 2. Generate Y by plugging X into $Y = F_Y^{-1}(X)$.
 $X \sim U(0, 1) \rightarrow F_Y^{-1}(\cdot) \rightarrow Y$ with CDF $F_Y(y)$

Pairs of Random Variables

Bivariate RV: $S_{X,Y} = \{(x,y) | \zeta \in S, X(\zeta) = x \text{ and } Y(\zeta) = y\}$

Joint CDF: $F_{X,Y}(x,y) = P[X \leq x, Y \leq y]$

Properties of joint CDF:

- a) $0 \leq F_{X,Y}(x,y) \leq 1$
- b) $F_X(x) = F_{X,Y}(x, \infty)$
- c) $F_Y(y) = F_{X,Y}(\infty, y)$
- d) $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$
- e) $F_{X,Y}(\infty, \infty) = P[X \leq \infty, Y \leq \infty] = 1$
- f) If $x_1 \leq x_2$ and $y_1 \leq y_2$, then $F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2)$
Means that $F_{X,Y}(x,y)$ is non-decreasing in the "northeast" direction

Joint PMF for discrete RVs X and Y : $p_{X,Y}(x,y) = P[X = x, Y = y]$

Properties of joint PMF:

- a) $0 \leq p_{X,Y}(x,y) \leq 1$ for all $(x,y) \in S_{X,Y}$
- b) $\sum \sum_{(x,y) \in S_{X,Y}} p_{X,Y}(x,y) = 1$
- c) When $(X,Y) \in B$, we say event B occurs and

$P[B] = P[(X,Y) \in B] = \sum \sum_{(x,y) \in B} p_{X,Y}(x,y)$

Marginal PMF:

$p_X(x) = \sum_{y \in S_Y} p_{X,Y}(x,y)$, $p_Y(y) = \sum_{x \in S_X} p_{X,Y}(x,y)$
 $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$, $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx$

Joint PDF: $F_{X,Y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u,v) du dv$, $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$

Some properties of joint PDF:

- a) $P[x_1 < X \leq x_2, y_1 < Y \leq y_2] = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1)$
- b) $f_{X,Y}(x,y) \geq 0$ for all (x,y)
- c) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1$
- d) $P[B] = \iint_{(x,y) \in B} f_{X,Y}(x,y) dx dy$

Function of two RVs, when $W = g(X,Y)$:

PMF: $p_W(w) = \sum \sum_{(x,y) | g(x,y)=w} p_{X,Y}(x,y)$

CDF: $F_W(w) = \int \int_{(x,y) \in \{g(x,y) \leq w\}} f_{X,Y}(x,y) dx dy$, $f_W(w) = \frac{dF_W(w)}{dw}$

CDF of $W = \max(X,Y)$: $F_W(w) = \int_{-\infty}^w \int_{-\infty}^w f_{X,Y}(x,y) dx dy$

Expected value (discrete): $E[W] = \sum_{x \in S_X} \sum_{y \in S_Y} g(x,y) p_{X,Y}(x,y)$

Expected value (continuous): $E[W] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$

Another theorem:

$E[g_1(X,Y) + \dots + g_n(X,Y)] = E[g_1(X,Y)] + \dots + E[g_n(X,Y)]$

Covariance of two RVs

Covariance: $Cov[X,Y] = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$

Variance rule thing: $Var[X + Y] = Var[X] + Var[Y] + 2Cov[X,Y]$

Correlation coefficient: $\rho_{X,Y} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$, it is always $-1 \leq \rho_{X,Y} \leq 1$

$\rho_{X,Y} = +1$ is strong positive relationship, -1 is strong negative, 0 is uncorrelated

Conditioning and Independent RVs

Conditional joint PMF: $p_{X,Y|B}(x,y) = \begin{cases} \frac{p_{X,Y}(x,y)}{P[B]}, & (x,y) \in B \\ 0, & \text{otherwise} \end{cases}$, $\sum \sum_{(x,y) \in B} p_{X,Y|B}(x,y) = 1$

Marginal PMFs: $p_{X|B}(x) = \sum_{y \in S_Y} p_{X,Y|B}(x,y)$, $p_{Y|B}(y) = \sum_{x \in S_X} p_{X,Y|B}(x,y)$

Conditional joint PDF: $f_{X,Y|B}(x,y) = \begin{cases} \frac{f_{X,Y}(x,y)}{P[B]}, & (x,y) \in B \\ 0, & \text{otherwise} \end{cases}$, $\int \int_{(x,y) \in B} f_{X,Y|B}(x,y) dx dy = 1$

Marginal PDFs: $f_{X|B}(x) = \int_{-\infty}^{\infty} f_{X,Y|B}(x,y) dy$, $f_{Y|B}(y) = \int_{-\infty}^{\infty} f_{X,Y|B}(x,y) dx$

Conditional exp. val. discrete: $E[W|B] = \mu_{W|B} = \sum \sum_{(x,y) \in B} g(x,y) p_{X,Y|B}(x,y)$

Conditional exp. val. continuous: $E[W|B] = \mu_{W|B} = \int \int_{(x,y) \in B} g(x,y) f_{X,Y|B}(x,y) dx dy$

Variance conditional: $Var[W|B] = \sigma_{W|B}^2 = E[W^2|B] - (E[W|B])^2$

Another conditional PMF: $p_{X,Y}(x,y) = p_{X|Y}(x|y)p_Y(y) = p_{Y|X}(y|x)p_X(x)$

Another conditional PDF: $f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x)$

Another expected value discrete: $E[g(X,Y)|Y = y] = \sum_{x \in S_X} g(x,y) p_{X|Y}(x|y)$

Another expected value continuous: $E[g(X,Y)|Y = y] = \int_{-\infty}^{\infty} g(x,y) f_{X|Y}(x|y) dx$

Independence, iff: $p_{X,Y}(x,y) = p_X(x)p_Y(y)$, $f_{X,Y}(x,y) = f_X(x)f_Y(y) = F_X(x)F_Y(y)$

Important stuff to remember about independence:

- a) $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$
- b) $E[XY] = E[X]E[Y]$
- c) $Cov[X,Y] = 0$ Note: iff = if and only if
- d) $Var[X + Y] = Var[X] + Var[Y]$
- e) $E[X|Y = y] = E[X]$, $E[Y|X = x] = E[Y]$

Bivariate Gaussian

Bivariate standard Gaussian: $f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(x^2 - 2\rho xy + y^2)}{2(1-\rho^2)}}$

Marginal PDFs: $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$

General bivariate gaussian RVs:

$$f_{X,Y}(x,y) = \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

Conditional (You can switch around X and Y , same thing):

Recall Gaussian PDF looks like: $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$

$f_{Y|X}(y|x) = \frac{1}{\tilde{\sigma}_2\sqrt{2\pi}} e^{-\frac{(y-\tilde{\mu}_2(x))^2}{2\tilde{\sigma}_2^2}}$ $\Rightarrow Y|X \sim N(\tilde{\mu}_2(x), \tilde{\sigma}_2^2)$

where $\tilde{\mu}_2(x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)$ and $\tilde{\sigma}_2^2 = \sigma_2^2(1 - \rho^2)$.

Consequently, given $X = x$, the conditional expected value and variance of Y are

$E[Y|X = x] = \tilde{\mu}_2(x)$ and $VAR[Y|X = x] = \tilde{\sigma}_2^2$

Bivariate Gaussian RVs are uncorrelated iff X and Y are independent

Miscellaneous miscellanea

PDF of $W = X + Y$: $f_W(w) = \int_{-\infty}^{\infty} f_{X,Y}(x, w - x) dx = \int_{-\infty}^{\infty} f_{X,Y}(w - y, y) dy$

When independent (same for y): $\int_{-\infty}^{\infty} f_{X,Y}(x, w - x) dx = \int_{-\infty}^{\infty} f_X(x) f_Y(w - x) dx$

Characteristic function (CHF): $\Phi_X(\omega) = E[e^{j\omega x}] = \int_{-\infty}^{\infty} e^{j\omega x} f_X(x) dx$

Inverse Fourier transform: $f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_X(\omega) e^{-j\omega x} d\omega$

Moment theorem: $\frac{1}{j^n} \frac{d^n}{d\omega^n} \Phi_X(\omega)|_{\omega=0}$

When $Z = X + Y$: $\Phi_Z(\omega) = \Phi_X(\omega)\Phi_Y(\omega)$

Central Limit Theorem, where $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$, also $E[Z_n] = 0$, $Var[Z_n] = 1$:

$$\lim_{n \rightarrow \infty} P[Z_n \leq z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx$$

Parameter estimation

Bias: $B = E[\hat{\theta}] - \theta$, unbiased if $B = 0$

Consistency: $\lim_{n \rightarrow \infty} P[|\hat{\theta}_n - \theta| > \epsilon] = 0$ for $\epsilon > 0$

Variance: $\sigma_{\hat{\theta}}^2 = E[(\hat{\theta} - E[\hat{\theta}])^2]$

Mean squared error: $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$

Efficiency rule: better if var is lower. Minimum var is MVUE

Minimum mean squared error: $MSE(\hat{\theta}_{MMSE}) \leq MSE(\hat{\theta}_0)$

Maximum likelihood estimation: $\hat{\theta}_{ML} = \max L(\theta)$ or $\max \ln L(\theta)$