Prob, Cons Prob & Counting

Sampling: n items k draws:

no replacement ordered: $n \cdot (n-1) \cdots (n-k+1)$ unordered: $\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$

replacement ordered: n^k unordered: $\binom{n+k-1}{k}$

Axioms of Prob: $P(\varnothing) = 0, P(S) = 1$ $\quad P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$ if disjoint

Inc-Exc: $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i P(A_i) - \sum_{i<j} (A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots$

Chain Rule: $P(A_1 \cap \cdots \cap A_n) = \prod_{j=1...k} P(A_j | A_1 \cap \cdots \cap A_{j-1})$

LOTP: $P(B) = \sum_{j=1}^{n} P(B|A_j)P(A_j)$ if $A_1, \ldots, A_n$ is a partition

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

Cond Indep: $P(A \cap B|C) = P(A|C)P(B|C)$

Cond Bayes: $P(A|B, E) = \frac{P(B|A, E)P(A|E)}{P(B|E)}$

Cond LOTP: $P(B|E) = P(B|A, E)P(A|E) + P(B|A^c, E)P(A^c|E)$

Random Variables & Joint Distribution

Indicator rv: $I(A) = 1$ if A occurs, 0 otherwise $\quad E(I(A)) = P(A)$

PMF Properties (drv/crv):

I. Nonnegative: $P(X = x) \geq 0 / f(x) \geq 0$

II. Sums/Integrates to 1: $\sum P(X = x) = 1 \int_{-\infty}^{\infty} f(x)dx = 1$

CDF Properties: $F(x) = \int_{-\infty}^{x} f(t)dt$ for continuous

I. Increasing: $x_1 \leq x_2 \to F(x_1) \leq F(x_2)$

II. Right continuous: $F(a) = \lim_{x \to a^+} F(x)$

III. Convergence: $\lim_{x \to -\infty} F(x) = 0 \wedge \lim_{x \to \infty} F(x) = 1$

Expectation:

Discrete: $E(X) = \sum xP(X = x)$ $\quad$ Continuous: $E(X) = \int_{-\infty}^{\infty} xf(x)dx$

Expectation Properties: (linearity)

$E(cX) = cE(X) \quad E(X+Y) = E(X) + E(Y)$

LOTUS:

Discrete: $E(g(x)) = \sum_x g(x)P(X = x)$

Discrete 2D: $E(g(X, Y)) \sum_x \sum_y g(x, y)P(X = x, Y = y)$

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

Continuous 2D: $E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f_{X,Y}(x, y)dxdy$

Variance: $\text{Var}(X) = E(X - \mu)^2 = E(X^2) - (E(X))^2$

Variance Properties:

I. $\text{Var}(cX) = c^2\text{Var}(X) \quad \text{Var}(X + c) = \text{Var}(X)$

II. $\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var(Y)}$ unless indep

Covariance: indep $\Rightarrow \text{Cov}(X, Y) = 0$

$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$

Covariance Properties:

$\text{Cov}(X, X) = \text{Var}(X) \quad \text{Cov}(X, Y) = \text{Cov}(Y, X) \quad \text{Cov}(X, c) = 0$

$\text{Cov}(X, Y + c) = \text{Cov}(X, Y) \quad \text{Cov}(aX, Y) = a\text{Cov}(X, Y)$

$\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$

$\text{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j)$

Correlation: $\text{Corr(X, Y)} = \frac{\text{Cov(X, Y)}}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \quad -1 \leq \text{Corr} \leq 1$

Named Distributions

- **BINOMIAL** $\text{Bin(n, p)} \equiv \sum n \text{ i.i.d Bern}(p)$

$X \sim \text{Bin}(n, p),\ Y \sim \text{Bin}(m, p) \Rightarrow X + Y \sim \text{Bin}(n + m, p)$

- **NEGATIVE BINOMIAL** $\text{NBin(r, p)} \equiv \sum r \text{ i.i.d Geom}(p)$
- **POISSON** $X \sim \text{Pois}(\lambda_1),\ Y \sim \text{Pois}(\lambda_2) \Rightarrow X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$
- **UNIFORM**

Universality of the Uniform: F a continuous & strictly increasing CDF

I. Let $U \sim \text{Unif}(0, 1)$, $X = F^{-1}(U)$, then X is an r.v. with CDF F

II. Let X be an r.v. with CDF F, then $F(X) \sim \text{Unif}(0, 1)$

- **NORMAL**

Symmetry of $\mathcal{Z}$:

I. Symmetry of PDF: $\varphi(z) = \varphi(-z)$

II. Symmetry of CDF tail areas: $\Phi(z) = 1 - \Phi(-z)$

III. Symmetry of $\mathcal{Z}$ and $-\mathcal{Z}$: $P(-\mathcal{Z} \leq z) = P(\mathcal{Z} \geq -z) = 1 - \Phi(-z)$

- **EXPO**

Memoryless: $P(X \geq s + t | X \geq s) = P(X \geq t)$, Expo the only memoryless c.r.v

$T_1 \sim \text{Expo}(\lambda_1),\ T_2 \sim \text{Expo}(\lambda_2) \Rightarrow P(T_1 < T_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$

- **BETA** $\text{Beta}(1, 1) \equiv \text{Unif}(0, 1)$

Normalising factor: $\beta(a, b) = \int_0^1 x^{a-1}(1 - x)^{b-1}dx$

$a < 1, b < 1 \Rightarrow$ U-shaped upward $\quad a > 1, b > 1 \Rightarrow$ U-shaped downward

$a = b \Rightarrow$ symmetric about $1/2$ $\quad a < | > b \Rightarrow$ favours < | > than $1/2$

- **GAMMA**

Gamma Function: $\Gamma(a) = \int_0^\infty x^a e^{-x} \frac{dx}{x}$

$\Gamma(a+1) = a\Gamma(a) \quad \Gamma(n) = (n-1)! \quad \Gamma(\frac{1}{2}) = \sqrt{\pi}$

$X \sim \text{Gamma}(a, \lambda), Y \sim \text{Gamma}(b, \lambda) \Rightarrow X + Y \sim \text{Gamma}(a+b, \lambda)$

$X \sim \text{Gamma}(a, 1) \Rightarrow E(X^c) = \frac{\Gamma(a+c)}{\Gamma(a)}$

$Y \sim \text{Gamma}(a, \lambda) \Rightarrow E(Y^c) = \frac{1}{\lambda^c} \frac{\Gamma(a+c)}{\Gamma(a)}$

Connections Between Distributions

Pois & Bin

$X \sim \text{Pois}(\lambda_1),\ Y \sim \text{Pois}(\lambda_2) \Rightarrow X|(X+Y=n) \sim \text{Bin}(n, \lambda_1/(\lambda_1+\lambda_2))$

$N \sim \text{Pois}(\lambda),\ X|(N=n) \sim \text{Bin}(n,p) \Rightarrow X \sim \text{Pois}(\lambda p),\ N - X \sim \text{Pois}(\lambda(1-p))$

X and $N - X$ indep

Pois & Expo

Poisson Process: $P(T_n > t) = P(N_t < n)$

I. #arrivals in interval t is $\text{Pois}(\lambda t)$

II. #arrivals in disjoint intervals are indep

$X_1 = t_1, X_2 = t_2 - t_1, \cdots \Rightarrow X_1, X_2, \cdots \overset{\text{i.i.d}}{\sim} \text{Expo}(\lambda)$

Gamma & Expo

$X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \text{Expo}(\lambda) \Rightarrow X_1 + \cdots + X_n \sim \text{Gamma}(n, \lambda)$

Beta & Gamma

$X \sim \text{Gamma}(a, \lambda), Y \sim \text{Gamma}(b, \lambda) \Rightarrow T = X + Y \sim \text{Gamma}(a+b, \lambda)$

$W = \frac{X}{X+Y} \sim \text{Beta}(a,b),\ \frac{1}{\beta(a,b)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$

T and W indep

Beta & Bin Conjugacy

Prior dist: $p \sim \text{Beta}(a,b),\ X|p \sim \text{Bin}(n,p)$

Posterior dist: $p|(X=k) \sim \text{Beta}(a+k, b+n-k)$

$P(X=k) = \binom{n}{k} \frac{\beta(a+k, b+n-k)}{\beta(a,b)}$

Gamma & Pois Conjugacy

Prior dist: $\lambda \sim \text{Gamma}(n_0, \lambda_0), X|\lambda \sim \text{Pois}(\lambda t)$

Posterior dist: $\lambda|(X=x) \sim \text{Gamma}(n_0 + x, \lambda_0 + t)$

Moments

Nth moment of X: $E(X^n)$ $\quad$ MGF: $M(t) = E(e^{tX})$

Discrete: $\sum_x e^{tx} P(X=x)$ $\quad$ Continuous: $\int_{-\infty}^{\infty} e^{tx} f_X(x) dx$

MGF is useful because:

I. Can be used to calculate moments

II. Determines a distribution uniquely

III. Works well with sum of indep r.v.s $M_{X+Y}(t) = M_X(t) M_Y(t)$

Transformations

1-var: $Y = g(X) \Rightarrow f_Y(y) = f_X(x) \left|\frac{dx}{dy}\right|,\ x = g^{-1}(y)$ $\qquad f_Y(y)dy = f_X(x)dx$

multi-var:

$$\vec{Y} = g\left(\vec{X}\right), \vec{X} = (X_1, \ldots, X_n) \Rightarrow f_{\vec{Y}}(\vec{y}) = f_{\vec{X}}(\vec{x}) \cdot \left|\det\left(\frac{\partial x}{\partial y}\right)\right|,\ x = g^{-1}(y)$$

Jacobian matrix: $\frac{\partial x}{\partial y} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix}$

Convolutions: determine dist of $T = X + Y$, X, Y indep: story/MGF/convolution

Discrete: $P(T=t) = \sum_x P(Y = t-x) P(X=x) = \sum_y P(X = t-y) P(Y=y)$

Continuous: $f_T(t) = \int_{-\infty}^{\infty} f_Y(t-x) f_X(x) dx = \int_{-\infty}^{\infty} f_X(t-y) f_Y(y) dy$

Multinomial

$\vec{X} \sim \text{Mult}_k(n, \vec{p}) \Rightarrow \sum_{j=1}^k p_j = 1,\ X_1 + \cdots + X_k = n$

PMF: $\frac{n!}{n_1! n_2! \cdots n_k!} \cdot p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$

Marginal: $X_j \sim \text{Bin}(n, p_j)$

Lumping: $(X_1 + X_2, X_3, \ldots, X_k) \sim \text{Mult}_{k-1}(n, (p_1 + p_2, p_3, \ldots, p_k))$

Cond: $(X_2, \ldots, X_k)|(X_1 = n_1) \sim \text{Mult}_{k-1}(n - n_1, (p_2', \ldots, p_k'))$

$P_j' = p_j / ((p_2 + \cdots + p_k))$

Cov: $i \neq j$, $\text{Cov}(X_i, X_j) = -np_ip_j$

Corr: $\text{Corr}(X_i)(X_j) = -\sqrt{\frac{p_ip_j}{(1-p_i)(1-p_j)}}$

Order Statistics

$P(X_{(j)} \le x) = P(\text{at least } j \text{ to the left of } x) = \sum_{k=j}^{n} \binom{n}{k} F(x)^k (1 - F(x))^{n-k}$

$f_{X_{(j)}}(x)dx = P(j-1 \text{ left}, n-j \text{ right}) = n \cdot \binom{n-1}{j-1} f(x)F(x)^{j-1}(1-F(x))^{n-j}$

$X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \text{Unif}(0,1) \Rightarrow X_{(j)} \sim \text{Beta}(j, n-j+1)$

$E(X_{(j)}) = \dfrac{j}{n+1}$

Cond Expectation & Variance

Cond Expectation Given Event:

Discrete: $E(Y|A) = \sum_y yP(Y=y|A)$ Continuous: $E(Y|A) = \int_{-\infty}^{\infty} yf(y|A)dy$

Cond Expectation Given r.v.: $g(x) = E(Y|X=x)$, $E(Y|X) = g(X)$

Properties:

I. Dropping what's independent: X, Y indep $\Rightarrow E(Y|X) = E(Y)$

II. Taking out what's known: $E(h(X)Y|X) = h(X)E(Y|X)$

III. Linearity: $E(Y_1 + Y_2) = E(Y_1|X) + E(Y_2|X), E(cY|X) = cE(Y|X)$

IV. **Adam's Law**: $E(E(Y|X)) = E(Y), E(E(Y|X,Z)|Z) = E(Y|Z)$

LOTE: $E(Y) = \sum_{j=1}^{n} E(Y|B_j)P(B_j)$ $\quad Y = I(B) \Rightarrow E(Y) = P(B)$ reduces to LOTP

Cond Variance:

$\text{Var}(Y|X) = E((Y - E(Y|X))^2|X) = E(Y^2|X) - (E(Y|X))^2$

Eve's law: $\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$ (within-grp + between-grp)

Inequalities & Limit Theorems

Markov: $P(|X| \ge a) \le E(|X|)/a$

Chebyshev: $P(|X - \mu| \ge a) \le \sigma^2/a^2$ $\quad P(|X-\mu| \ge c\sigma) = 1/c^2$

Cauchy-Schawarz: $|E(XY)| \le \sqrt{E(X^2)E(Y^2)}$

Jensen: g convex $\Rightarrow E(g(X)) \ge g(E(X))$ g concave $\Rightarrow E(g(X)) \le g(E(X))$

Sample Mean & Variance:

$\bar{X}_n = (X_1 + \cdots + X_n)/n \qquad E(\bar{X}_n) = \mu \qquad \text{Var}(\bar{X}_n) = \sigma^2/n$

$S_n^2 = 1/(n-1)\sum_{j=1}^{n}(X_j - \bar{X}_n)^2$

Strong LLN: $P(\bar{X}_n \to \mu) = 1 \text{ as } n \to \infty$

Weak LLN: $\forall \epsilon > 0, P(|\bar{X}_n - \mu| \ge \epsilon) \to 0 \text{ as } n \to \infty$

CLT: $\sqrt{n}\left(\dfrac{\bar{X}_n - \mu}{\sigma}\right) \to \mathcal{N}(0,1) \text{ as } n \to \infty$

PDF/PMF	**Discrete**	**Continuous**
Joint	$p_{X,Y}(x,y) = P(X=x, Y=y)$ $P((X,Y) \in A) = \sum\sum_{(x,y)\in A} P(X=x, Y=y)$	$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$ $P((X,Y) \in A) = \iint_A f_{X,Y}(x,y)dxdy$
Marginal	$P(X=x) = \sum_y P(X=x, Y=y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$
Cond	$P(Y=y \mid X=x) = \frac{P(X=x, Y=y)}{P(X=x)}$	$f_{Y\mid X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$
Indep	$P(X=x, Y=y) = P(X=x)P(Y=y)$	$f_{X,Y}(x,y) = f_X(x) f_Y((y))$

Discrete $X \sim$	**PMF** $P(X=k)$	$E(X)$	$\mathrm{Var}(X)$	**MGF** $M(t)$
$\mathrm{Bern}(p)$ One time	$k=1 \Rightarrow p$ $k=0 \Rightarrow 1-p$	p	$p(1-p)$	$pe^t + (1-p)$
$\mathrm{Bin}(n,p)$ #S in n trials	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$	$(pe^t + (1-p))^n$
$\mathrm{Geom}(p)$ #F before 1^{st} S	$p(1-p)^k$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	$\frac{p}{1-(1-p)e^t}$
$\mathrm{NBin}(r,p)$ #F before r^{th} S	$\binom{k+r-1}{r-1} p^r (1-p)^k$	$r\frac{1-p}{p}$	$r\frac{1-p}{p^2}$	$\left(\frac{p}{1-(1-p)e^t}\right)^r$
$\mathrm{HGeom}(w,b,n)$ #w in chosen n	$\frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$	$n\frac{w}{w+b}$	$\frac{w+b-n}{w+b-1} np(1-p)$	$\times$
$\mathrm{Pois}(\lambda)$	$\frac{e^{-\lambda}\lambda^k}{k!}$	λ	λ	$\exp\left(\lambda(e^t-1)\right)$

Continuous	**PDF** $f(x)$	**CDF** $F(x)$	$E(X)$	$\mathrm{Var}(X)$	**MGF** $M(t)$
$U \sim \mathrm{Unif}(a,b)$	$\frac{1}{b-a}$	$\frac{x-a}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
$\mathcal{Z} \sim \mathcal{N}(0,1)$	$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$	$\Phi(z)$	0	1	$e^{t^2/2}$
$X \sim \mathcal{N}(\mu, \sigma^2)$ $X = \mu + \sigma\mathcal{Z}$	$\varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}$	$\Phi\left(\frac{x-\mu}{\sigma}\right)$	μ	σ^2	$e^{\mu t + \frac{1}{2}\sigma^2 t^2}$
$X \sim \mathrm{Expo}(\lambda)$ $\lambda X \sim \mathrm{Expo}(1)$	$e^{-\lambda x}$	$1 - e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - t}$
$X \sim \mathrm{Gamma}(a, \lambda)$ $\lambda X \sim \mathrm{Gamma}(a, 1)$	$\frac{1}{\Gamma(a)} (\lambda x)^a e^{-\lambda x} \frac{1}{x}$	$\times$	$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$	$\left(\frac{\lambda}{\lambda - t}\right)^a$
$X \sim \mathrm{Beta}(a,b)$	$\frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1}$	$\times$	$\frac{a}{a+b}$	$\times$	$\times$