Introduction to Probability

Sample Space and Probability

$$P(A\cup B)=P(A)+P(B)-P(A\cap B) \quad \text{(Venn diagram)}$$

$$P(\cup_{i=1}^{n}A_i)\le\sum_{i=1}^{n}P(A_i) \quad \text{(union bound)}$$

$$P(B)=P(A_1)P(B|A_1)+\cdots P(A_n)P(B|A_n) \quad \text{(total probability)}$$

$$P(\cap_{i=1}^{n}A_i)=P(A_1)P(A_2|A_1)\cdots P(A_n|\cap_{i=1}^{n}A_i) \quad \text{(chain rule)}$$

$$P(A_i|B)=\frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n}P(B|A_i)P(A_i)} \quad \text{(Bayes rule)}$$

$$P(\cap_{i\in S}A_i)=\prod_{i\in S}P(A_i) \quad \text{for every subset S of } \{1,2,...,n\} \quad \text{(indep.)}$$

$$E[X]=\sum_{x\in X}xp_x(x), \text{VAR}(X)=E[(X-E[X])^2]=E[X^2]-E[X]^2$$

$$p_x(x)=\sum_y p_{X,Y}(x,y), p_y(y)=\sum_x p_{X,Y}(x,y) \quad \text{(marginals)}$$

$$p_Y(y)=\sum_{\{x|g(x)=y\}}p_X(x) \quad \text{(functions of RV)}$$

$$E[X|Y=y]=\sum_x xp_{X|Y}(x|y) \quad \text{(conditional expectation)}$$

$$E[X]=E[E[X|Y]] \quad \text{(iterated expectation)}$$

$$\text{VAR}(X)=E[\text{VAR}(X|Y)]+\text{VAR}(E[X|Y]) \quad \text{(law of total variance)}$$

$$\text{VAR}(\sum_{i=1}^{n}X_i)=\sum_{i=1}^{n}\text{VAR}(X_i)+\sum_{i\neq j}\text{COV}(X_i,X_j) \quad \text{(variance of sum)}$$

Discrete Random Variables

Bernoulli: X = {heads of a biased coin}
$\text{p}_x(x;p)=p^x(1-p)^{1-x}, x\in\{0,1\}$
$E[X]=p,\ \text{VAR}(x)=p(1-p)$

Binomial: X = {number of heads (successes) in n trials}
$\text{p}_x(x;p)=\binom{n}{k}p^k(1-p)^{n-k}, k=0,1,...,n$
$S_n=\sum_{i=1}^{n}X_i \sim \text{Bin}(n,k;p), \text{where} X_i\sim\text{Ber}(p)$
$E[S_n]=np,\ \text{VAR}(S_n)=np(1-p)$

Geometric: X = {number of trials until the first success}
$\text{p}_x(x;k)=(1-p)^{k-1}p, k=1,2,...$
$E[X]=1/p, \text{VAR}(x)=\frac{1-p}{p^2}$

Poisson: X = {number of arrivals}
$\text{p}_x(k;\lambda)=\frac{\lambda^k}{k!}\exp(-\lambda), k=0,1,2,...$
$E[X]=\lambda, \text{VAR}(x)=\lambda$

Continuous Random Variables

Uniform: X = {equally likely}
$\text{f}_x(x)=1/(b-a), \quad \text{if } a\le x\le b$
$E[X]=(a+b)/2,\ \text{VAR}(x)=(b-a)^2/12$

Exponential: X = {lifetime duration}
$\text{f}_x(x)=\lambda\exp\{-\lambda x\}, \quad \text{if } x\ge 0$
$E[X]=(a+b)/2,\ \text{VAR}(x)=(b-a)^2/12$

Poisson: X = {event arrivals}
$\text{f}_x(x;\lambda,\tau)=\frac{(\lambda\tau)^k}{k!}\exp\{-\lambda\tau\}$
$E[X]=\lambda\tau,\ \text{VAR}(x)=\lambda\tau$

Gaussian: X = {law of large numbers}
$\text{f}_x(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$
$f_x(x;\mu,\Sigma)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$
$E[X]=\mu,\ \text{VAR}(X)=\Sigma$

Further Topics on Random Variables

Let $Y=g(X)$ a transformation of rv X then:

$$F_Y(y)=P(g(X)\le y)=P(X\le g^{-1}(y))=F_X(g^{-1}(y))$$

If g(x) is one-to-one, we have:

$$f_Y(y)=f_X(x)|\frac{dx}{dy}|=f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)|$$

Sum of indep. rvs X and Y is the convolution of their PDFs:

$$p_Z(z)=P(X+Y=z)=\sum_{\{(x,y):x+y=z\}}P(X=x,Y=y)$$

$$=\sum_x P(X=x,Y=z-x)=\sum_x p_X(x)p_Y(z-x)$$

Covariance and correlation:

$$\text{COV}(X,X)=\text{VAR}(X)=\sigma_X^2$$

$$\text{COV}(X,aY+b)=a\text{COV}(X,Y)$$

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

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

$$\text{VAR}(X+Y)=\text{VAR}(X)+\text{VAR}(Y)+2\text{COV}(X,Y)$$

$$\rho(X,Y)=\frac{\text{COV}(X,Y)}{\sigma_X\sigma_Y}\in[-1,1]$$

Moment Generating Function:

$$M_X(s)=E[e^{sX}]=\sum_x p_X(x)e^{sx}; E[X^n]=\frac{d^n}{ds^n}M_X(s)|_{s=0}$$

Further Topics on Random Variables

Sum of a random number of RVs:
Let $Y=X_1+...+X_N$, where N is a RV

$$E[Y|N=n]=E[X_1+...+X_n|N=n]=nE[X]\to E[Y|N]=NE[X]$$

$$E[Y]=E[E[Y|N]]=E[NE[X]]=E[N]E[X]$$

$$\text{VAR}(Y|N=n)=\text{VAR}(X_1+...+X_N|N=n)=n\text{VAR}(X)$$

$$\text{VAR}(Y)=E[\text{VAR}(Y|N)]+\text{VAR}(E[Y|N])$$

$$=E[N\text{VAR}(X)]+\text{VAR}(NE[X])=\text{VAR}(X)E[N]+(E[X])^2\text{VAR}(N)$$

Inequalities and Limit Theorems

$$P(X\ge a)\le\frac{E[X]}{a}, \quad \text{for } a>0 \quad \text{(Markov inequality)}$$

$$P(|X-\mu|\ge\epsilon)\le\frac{\sigma^2}{\epsilon^2}, \quad \text{for } \epsilon>0 \quad \text{(Chebyshev inequality)}$$

$$P(X\ge a)=P(e^{sX}\ge e^{sa})\le E[e^{sX}]e^{-sa} \quad \text{(Chernoff inequality)}$$

WLLN: Let $M_n=\frac{1}{n}(X_1+...+X_n)$, where X_i are iid, then:
$M_n\to\mu$ in prob., i.e. $\lim_{n\to\infty}P(|M_n-\mu|>\epsilon)=0$

SLLN: X_i are iid, then: $P(\lim_{n\to\infty}(\frac{1}{n}(X_1+...+X_n)=\mu)=1$

CLT: Let $S_n=(X_1+...+X_n)=nM_n$ where X_i are iid, then:

$$Z_n=\frac{S_n-n\mu}{\sigma\sqrt{n}}\sim N(0,1), \text{i.e.} \lim_{n\to\infty}P(Z_n\le z)=\Phi(z)$$

Let ϵ be accuracy and δ be confidence level, then:
$P(|Y_n-a|\ge\epsilon)\le\delta \quad \forall\ n\ge n_0$

Bayesian Inference

MAP: $\hat{\theta}_{MAP}=\arg\max_\theta p(\theta|x)=\arg\max_\theta p(x|\theta)p(\theta)$

LMS: $\hat{\theta}_{LMS}=E_{P(\theta|x)}[\theta|X=x]=f(x)$

Hypothesis Testing: $P(\theta=\theta_1|X=x)\gtrless_{H2}^{H1}P(\theta=\theta_2|X=x)$

L-LMS: $\hat{\theta}=E[\theta]+\frac{\text{COV}(\theta,X)}{\text{VAR}(X)}(X-E[X])$

E.g. $X_i=\theta+W_i,\ \theta\sim N(\mu,\sigma_0^2), W_i\sim N(0,\sigma_i^2)$

$$\hat{\theta}_{LLMS}=\frac{\mu/\sigma_0^2+\sum_{i=1}^{n}x_i/\sigma_i^2}{\sum_{i=0}^{n}1/\sigma_i^2}$$

Bias: $b(\hat{\theta})=E[\hat{\theta}]-\theta$

Consistent: $\hat{\theta_n}\to\theta$ in prob, i.e. $\lim_{n\to\infty}P(|\hat{\theta_n}-\theta|>\epsilon)=0$

MSE: $E[(\hat{\theta_n}-\theta)^2]=(E[\hat{\theta}]-\theta)^2+\text{VAR}(\hat{\theta})=\text{bias}^2+\text{variance}$

Bernoulli Process

Let $X_1, X_2, ...$ be a sequence of iid Bernoulli(p)

Then probability of k arrivals in n time-steps:

$$p_N(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Let T_i be event inter-arrival times

$$T_i \sim \text{Geom}(p) = (1-p)^{k-1}p, \ k = 1, 2, ...$$

Let $Y_k = T_1 + T_2 ... + T_k$ be the total time then

$$E[Y_k] = kE[T] = \frac{k}{p}, \ \text{VAR}(Y_k) = k\text{VAR}(T) = \frac{k(1-p)}{p^2}$$

$$P_{Y_k}(t) = \binom{t-1}{k-1} p^k (1-p)^{t-k}, \ t = k, k+1, ... \text{(Pascal of order k)}$$

$$P(T = t + n | T > n) = P(T = T) = (1-p)^{t-1}p \text{ (memoryless)}$$

Poisson Process

Time-homogeneity: $P(k, \tau)$ prob of k arrivals is the same for same τ

Independence: Number of arrivals in disjoint intervals is independent

$$P(k, \tau) \text{ satisfy}: \ P(0, \tau) \approx 1 - \lambda\tau, \ P(1, \tau) \approx \lambda\tau, \ P(>1, \tau) \approx 0$$

$$P(k, \tau) = \frac{(\lambda\tau)^k}{k!} \exp\{-\lambda\tau\}, \text{ where } E[N_\tau] = \lambda\tau, \ \text{VAR}(N_\tau) = \lambda\tau$$

Let T_i be event inter-arrival times $\sim \text{Exp}(\lambda)$

$$f_T(t) = \lambda e^{-\lambda t}, t \geq 0; \ E[T] = \frac{1}{\lambda}; \ \text{VAR}(T) = \frac{1}{\lambda^2}$$

Let $Y_k = T_1 + T_2 ... + T_k$ be the total time then

$$E[Y_k] = kE[T] = \frac{k}{\lambda}, \ \text{VAR}(Y_k) = k\text{VAR}(T) = \frac{k}{\lambda^2}$$

$$f_{Y_k}(y) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}, \ y \geq 0 \text{ (Erlang pdf of order k)}$$

Markov Chain

$$r_{ij}(n) = P(x_n = j | x_0 = i)$$

$$r_{ij}(n) = \sum_{k=1}^{m} r_{ik}(n-1) p_{kj}$$

Recurrent class R is aperiodic iff there exists n s.t.

$$r_{ij}(n) > 0 \ \forall i, j \in \mathbb{R}$$

$\pi_j = P(X_k = j)$ when n is large (steady-state prob of j)

$\lim_{n \to \infty} r_{ij}(n) = \pi_j$ regardless of where you start

for recurrent, aperiodic, irreducible chain (1 class)

$\pi = \pi P$ where π is a row vector

$\pi_j > 0$ for all recurrent states

$\pi_j = 0$ for all transient states