Computational Probability - Definitions

1 Common Discrete Random Variables

Bernoulli Random Variable: A Bernoulli random variable, X, on the set $\Omega = \{0, 1\}$ is such that $P(X = 0) = p$ and $P(X = 1) = 1 - p$. Thus $E(X) = 1 - p$ and $Var(X) = p(1 - p)$.

Binomial Random Variable: A Binomial random variable, X, with parameters (n, k) on the set $\Omega = \{0, 1, \ldots, n\}$ is such that $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$. Thus $E(X) = np$ and $Var(X) = np(1 - p)$.

Geometric Random Variable: A Geometric random variable, X, with parameter p on the set $\Omega = \{1, 2, \ldots\}$ is such that $P(X = k) = p(1 - p)^{k-1}$. Thus $E(X) = 1/p$ and $Var(X) = (1 - p)/p^2$.

Poisson Random Variable: A Poisson random variable, X, with parameter λ on the set $\Omega = \{0, 1, 2, \ldots\}$ is such that $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$. Thus $E(X) = Var(X) = \lambda$.

2 Common Continuous Random Variables

Uniform Random Variable: A Uniform random variable, X, on the set $[a, b]$ is such that $f_X(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$. Thus $E(X) = \frac{1}{2}(a+b)$ and $Var(X) = (b-a)^2/12$.

Exponential Random Variable: A Uniform random variable, X, on the set $[0, \infty)$ with parameter $\lambda > 0$ is such that $f_X(x) = \lambda e^{-\lambda x}$. Thus $E(X) = 1/\lambda$ and $Var(X) = 1/\lambda^2$.

Gaussian (Normal) Random Variable: A Gaussian random variable on $\mathbb{R}$ with parameters (m, σ) is such that $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/(2\sigma^2)}$. Thus $E(X) = m$ and $Var(X) = \sigma^2$.

Gamma Random Variable: A Gamma random variable on $(0, \infty)$ with parameters (α, λ) is such that $f_X(x) = \frac{\lambda e^{-\lambda x} (\lambda x)^{\alpha - 1}}{\Gamma(\alpha)} \mathbf{1}_{(0,\infty)}(x)$. Thus $E(X) = \alpha/\lambda$ and $Var(X) = \alpha/\lambda^2$.

3 Counting

Permutations: The number of permutations of n objects is $n!$. If $n = \sum_{i=1}^{r} n_i$, where n_i are alike, then there are

$$\frac{n!}{\prod_{i=1}^{r} n_i!}$$

permutations. The number of permutations of n objects taken k at a time is $n!/(n-k)!$.

Combinations: We define $\binom{n}{k}$, for $k \leq n$, by

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

and say that $\binom{n}{k}$ represents the number of possible combinations of n objects taken k at a time (different orderings are not counted).

4 Jointly Distributed Random Variables

Joint Cumulative Distribution: The joint cumulative distribution of random variables (X, Y) is $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$. Then the marginal cumulation distributions of X and Y are

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y) \quad \text{and} \quad F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x, y).$$

Joint Probability Mass Function (pmf): The joint probability mass function of (discrete) random variables (X, Y) is defined an $f_{X,Y}(x, y) = P(X = x, Y = y)$. Then $P((X, Y) \in A) = \sum\sum_{(x_k, y_l) \in A} f_{X,Y}(x_k, y_l)$.

Joint Probability Density Function (pdf): The joint probability density function of (continuous) random variables (X, Y) is defined as $f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$. Then $P((X, Y) \in A) = \int\int_A f_{X,Y}(x, y)\, dx\, dy$.

Independence: Random variables X and Y are independent iff where exists functions $g(x)$ and $h(y)$ such that $f_{X,Y}(x, y) = g(x)h(y)$. Note that if we normalize $g(x)$ and $h(y)$ by their L^1 norms, then they are equal to the pdf's of X and Y, respectively.

Conditional Probability: Let X and Y be random variables. Then we define the conditional distribution of Y given $X = x$ by $F_{Y|X}(y|x) = P(Y \leq y | X = x)$.
If X is a discrete random variable, then

$$F_{Y|X}(y|x) = \frac{P(Y \leq y, X = x)}{P(X = x)} \quad \text{and} \quad f_{Y|X}(y|x) = \frac{P(X = x, Y = y)}{P(X = x)}.$$

If X is a continuous random variable, then

$$F_{Y|X}(y|x) = \frac{\int_{-\infty}^{y} f_{X,Y}(x,y')\,dy'}{f_X(x)} \quad \text{and} \quad f_{Y|X}(y|x) = \frac{\partial}{\partial y} F_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

Conditional Expectation: The conditional expectation of Y given $X = x$ is defined by

$$E(Y|X=x) = \int_{\mathbb{R}} y f_{Y|X}(y|x)\,dy.$$

Note that $E(Y|X=x)$ is a function of x and $E(Y) = E(E(Y|X))$.

Bayes Theorem: Then we have that for discrete and continuous random variables, respectively,

$$\begin{aligned} P(Y=y) &= \sum_{x_k} P(Y=y|X=x_k)P(X=x_k) \\ P(Y \in A) &= \int_{\mathbb{R}} P(Y \in A|X=x) f_X(x)\,dx. \end{aligned}$$

Sums of Independent Random Variables: Let X and Y be independent random variables. Then $f_{X+Y}(a) = f_X * f_Y(a)$.

Sums of Independent Gaussian Random Variables: Let $\{X_k\}$ be a sequence of independent Gaussian random variables with mean m_k and variance σ_k^2. Let $S_n = \sum_{k=1}^n X_k$. Then S_n is a Gaussian random variable with mean $\sum_{k=1}^n m_k$ and variance $\sum_{k=1}^n \sigma_k^2$.

Sums of Independent Exponential Random Variables: Let $\{X_k\}$ be a sequence of independent exponential random variables, with parameter λ. Let $S_n = \sum_{k=1}^n X_k$. Then S_n is a Erlang random variable with pdf

$$f_{S_n}(y) = \frac{(\lambda y)^{n-1}}{(n-1)!}\lambda e^{-\lambda y}, \quad y \geq 0.$$

Distributions of a Function of a Random Variable: Let $Y = g(X)$. Let $\{x_1, \ldots, x_n\}$ be such that $y = g(x_k)$. Then

$$f_Y(y) = \sum_{k=1}^n f_X(x) \left|\frac{dx}{dy}\right|\Bigg|_{x=x_k}.$$

Transformations of Random Variables: Let $\mathbf{X} = (X_1, \ldots, X_n)$ and $Z_k = g_k(\mathbf{X})$ for $k = 1, \ldots, n$. Then

$$\begin{aligned} F_{Z_1,\ldots,Z_n}(z_1,\ldots,z_n) &= P(g_1(\mathbf{X}) \leq z_1, \ldots, g_n(\mathbf{X}) \leq z_n) \\ &= \int_{\{\mathbf{x}': g_k(\mathbf{x}') \leq z_k\}} \cdots \int f_{X_1,\ldots,X_n}(x_1',\ldots,x_n')\,dx_1 \cdots dx_n. \end{aligned}$$

The next theorem provides a nice way to compute the pdf of a transformation.

Joint Probability Distribution of Functions if Random Variables: Let

$$\begin{aligned} Y_1 &= g_1(X_1,X_2), \quad Y_2 = g_2(X_1,X_2) \\ X_1 &= h_1(Y_1,Y_2), \quad X_2 = h_2(Y_1,Y_2) \end{aligned}$$

and

$$\begin{aligned} J_X^Y &= \det\left(\frac{\partial(Y_1,Y_2)}{\partial(X_1,X_2)}\right) = \det\left(\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{bmatrix}\right) \\ J_Y^X &= \det\left(\frac{\partial(X_1,X_2)}{\partial(Y_1,Y_2)}\right) = \det\left(\begin{bmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} \end{bmatrix}\right). \end{aligned}$$

Note that $J_X^Y = (J_Y^X)^{-1}$. Then

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= f_{Y_1,Y_2}(y_1,y_2)J_X^Y, \quad \text{where} \quad y_1 = g_1(x_1,x_2), y_2 = g_2(x_1,x_2) \\ f_{Y_1,Y_2}(y_1,y_2) &= f_{X_1,X_2}(x_1,x_2)J_Y^X, \quad \text{where} \quad x_1 = h_1(y_1,y_2), x_2 = h_2(y_1,y_2). \end{aligned}$$

Now if $A = (h_1,h_2)(B)$, then

$$\begin{aligned} \int_A\int f_{X_1,X_2}(x_1,x_2)\,dx_1\,dx_2 &= \int_B\int f_{Y_1,Y_2}(y_1,y_2)\,dy_1\,dy_2 \\ &= \int_B\int f_{X_1,X_2}(h_1(y_1,y_2),h_2(y_1,y_2))J_Y^X\,dy_1\,dy_2. \end{aligned}$$

One Function of Several Random Variables: Let $Z = g(X_1,\ldots,X_n)$ and $R_z = \{\mathbf{x} = (x_1,\ldots,x_n) : g(\mathbf{x}) \le z\}$. Then

$$F_Z(z) = P(Z \le z) = P(\mathbf{X} \in R_z) = \int_{\mathbf{x}\in R_z}\cdots\int f_{X_1,\ldots,X_n}(x_1',\ldots,x_n')\,dx_1'\ldots dx_n'.$$

Correlation and Covariance: Let X and Y be random variables. Then the covariance, correlation, and correlation coefficient of X and Y are respectively defined as

$$\begin{aligned} \mathrm{Cov}(X,Y) &= E[(X-E(X))(Y-E(Y))] = E(XY) - E(X)E(Y) \\ \mathrm{R}_{XY} &= E[XY] \\ \rho(X,Y) &= \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}. \end{aligned}$$

If $\mathrm{Cov}(X,Y) = 0$, then X and Y are said to be uncorrelated. Note that $\mathrm{Cov}(X,Y) = 0$ is equivalent to $E(XY) = E(X)E(Y)$. If $\mathrm{Cov}(X,Y) = 0$, we say that X and Y are orthogonal.

Joint Gaussian Random Variables: The random variables $X_1, X_2, \ldots, X_n$ are said to be jointly Gaussian if their joint pdf is given by

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{1}{(2\pi)^{n/2}|K|^{1/2}}e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T K^{-1}(\mathbf{x}-\mathbf{m})},$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{m} = \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix} = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{bmatrix},$$

and $(K)_{ij} = \mathrm{Cov}(X_i,X_j)$. The matrix K is called the covariance matrix.

Linear Transformation of Jointly Gaussian Random Variables: Let $X_1,\ldots,X_n$ be jointly Gaussian random variables and let $\mathbf{Y} = (Y_1,\ldots,Y_n)^T$ be defined by

$\mathbf{Y} = A\mathbf{X}$, where A is an invertible n by n matrix. Let $\mathbf{m}_y = E(\mathbf{Y})$ and $C = AKA^T$, where K is the correlation matrix of X_i. Then $\mathbf{Y}$ is jointly Gaussian and

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\sqrt{2\pi}|C|^{1/2}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{m}_y)^T C^{-1}(\mathbf{y}-\mathbf{m}_y)}.$$

Note that $\mathbf{m}_y = A\mathbf{m}_x$.

Proposition: Let X and Y be Gaussian random variables. Then X and Y are independent if and only if X and Y are uncorrelated.

5 Random Processes

Sample Mean and Variance: The sample mean and variance are given respectively by

$$\begin{aligned} \mu_n &= \frac{1}{n}\sum_{k=1}^{n} X_n \\ \sigma_n^2 &= \frac{1}{n}\sum_{k=1}^{n} (X_n - \mu_{n-1})^2. \end{aligned}$$

Independent Increments: A random process $X(t)$ is said to have independent increments if for $t_1 < t_2 < \cdots < t_k$, the random variables

$$X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_k) - X(t_{k-1})$$

are independent random variables.

Independent Random Processes: The random processes $X(t)$ and $Y(t)$ are said to be independent if the vector random variables $(X(t_1), \ldots, X(t_k))$ and $(Y(t_1'), \ldots, Y(t_j'))$ are independent for all k, j and for any $t_1, \ldots, t_k$ and $t_1', \ldots, t_j'$.

Markov Process: A random process is said to be Markov if for any k, $t_1 < t_2 < \cdots < t_k$, and for any $x_1, x_2, \ldots, x_k$ we have that

$$f_{X(t_k)}(x_k|X(t_{k-1}) = x_{k-1}, \ldots, X(t_1) = x_1) = f_{X(t_k)}(x_k|X(t_{k-1}) = x_{k-1})$$

if $X(t)$ is continuous-valued, and

$$P(X(t_k) = x_k|X(t_{k-1}) = x_{k-1}, \ldots, X(t_1) = x_1) = P(X(t_k) = x_k|X(t_{k-1}) = x_{k-1})$$

if $X(t)$ is discrete-valued.

First and Second Order Statistics: The first and second order statistics of the random process $X(t)$ are

Name	Formula
Mean	$m_X(t) = E[X(t)]$
Variance	$\text{Var}[X(t)] = E[(X(t) - m_X(t))^2]$
Autocorrelation	$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)]$
Autocovariance	$C_{XX}(t_1, t_2) = E[(X(t_1) - m_X(t_1))(X(t_2) - m_X(t_2))]$

Note that $\text{Var}[X(t)] = C_{XX}(t, t)$ and $C_{XX}(t_1, t_2) = R_{XX}(t_1, t_2) - m_X(t_1)m_X(t_2)$.

Second Order Statistics of Multiple Random Processes: The second order statistics of the random processes $X(t)$ and $Y(t)$ are

Name	Formula
Cross-Correlation	$R_{X,Y}(t_1, t_2) = E[X(t_1)Y(t_2)]$
Cross-Covariance	$C_{X,Y}(t_1, t_2) = E[(X(t_1) - m_X(t_1))(Y(t_2) - m_Y(t_2))]$

Stationary Increments: The random process S_n has stationary increments if

$$P(S_{n'} - S_n = y) = P(S_{n'-n} = y).$$

Poisson Process: The poisson process is given by $N(t)$ represents the number of event occurrences in the interval $[0, t]$ and its distribution is given by

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \ldots.$$

Note that $N(t)$ has both stationary and independent increments.

Distribution of Inter-arrival Time in a Poisson Process: Let T be a random variable that equals the time between event occurrences in a Poisson process. Then $P(T \leq t) = 1 - e^{-\lambda t}$, i.e., T is an exponentially distributed random variable.

Nth Order Stationary Random Process: A random process $X(t)$ is nth order stationary if

$$F_{X(t_1), X(t_2), \ldots, X(t_n)}(x_1, \ldots, x_n) = F_{X(t_1+\tau), X(t_2+\tau), \ldots, X(t_n+\tau)}(x_1, \ldots, x_n).$$

Stationary Random Processes: A random process $X(t)$ is stationary if it is stationary for all orders. In other words

$$F_{X(t_1), \ldots, X(t_k)}(x_1, \ldots, x_k) = F_{X(t_1+\tau), \ldots, X(t_k+\tau)}(x_1, \ldots, x_k),$$

for all $\tau \in \mathbb{R}$, all $k \in \mathbb{N}$, and all choices of time samples $t_1, \ldots, t_k$.

Cyclostationary: A random process $X(t)$ is said to be cyclostationary if the joint cdf function of any set of samples is invariant with respect to shifts of the origin by integer multiples of some period T. In other words,

$$F_{X(t_1), \ldots, X(t_k)}(x_1, \ldots, x_k) = F_{X(t_1+mT), \ldots, X(t_k+mT)}(x_1, \ldots, x_k).$$

Wide-Sense Stationary: A random process $X(t)$ is wide-sense stationary if

$$\begin{aligned} m_X(t) &= m \\ C_{XX}(t_1, t_2) &= C_{XX}(t_1 - t_2). \end{aligned}$$

Note that stationary processes are wide-sense stationary, but the converse is not necessarily true.

Jointly Wide-Sense Stationary: The random processes $X(t)$ and $Y(t)$ are jointly wide-sense stationary if they are both wide-sense stationary and if their cross-covariance depends only on $t_2 - t_1$.

Properties of the Autocorrelation of a Wide-Sense Stationary Random Process: We have $R_{XX}(\tau) = R_{XX}(-\tau)$ and $|R_{XX}(\tau)| \leq R_{XX}(0)$.

Time Average: The time average of a random process $X(t)$ is

$$< X(t) >_T = \frac{1}{2T} \int_{-T}^{T} X(t)\, dt.$$

6 Processing of Random Signals

Power Spectral Density (PSD): The power spectral density of $X(t)$, where $X(t)$ is a WSS random process is given by the Fourier transform of the autocorrelation function:

$$S_X(f) = \mathcal{F}[R_{XX}(\tau)] = \int_{\mathbb{R}} R_{XX}(\tau) e^{-j2\pi f \tau} d\tau.$$

Note that since $R_{XX}(\tau)$ is real and even, $S_X(f)$ is real, even, and nonnegative. Also note that it does not make sense to define the power spectral density for a random process that is not WSS.

Cross-Power Spectral Density: The cross-power spectral density of $X(t)$ and $Y(t)$, where $X(t)$ and $Y(t)$ are jointly WSS is given by

$$S_{X,Y}(f) = \mathcal{F}[R_{X,Y}(\tau)].$$

Note that $S_{XY}(f)$ is, in general, complex-valued.

Average Power: The average power of a random process $X(t)$ is

$$E[X^2(t)] = R_{XX}(0) = \int_{\mathbb{R}} S_X(f)\, df.$$

PSD of a Linear System: Suppose that $Y(t) = h(t) * X(t)$. Then

$$\begin{aligned} S_Y(f) &= |H(f)|^2 S_X(f) \\ S_{Y,X}(f) &= H(f) S_X(f) \\ S_{X,Y}(f) &= S^*_{Y,X}(f) = H^*(f) S_X(f) \end{aligned}$$