Probability and Statistics Notes

Prabath Peiris

Contents

1 Definitions

1.1 Sample Space ($\mathcal{S}$)

The set of all possible outcomes is the sample space $\mathcal{S}$.

1.2 Events

Events are sets of outcomes.

1.3 Random Variable

Random Variable is a function that maps the sample space $\mathcal{S}$ into a subset of the real line.

1.4 Probability Mass Function (pmf)

probability distribution or *probability mass function* of a discrete random variable is define for every number x by $p(x)$

2 Lecture 2: (Chapter 4) Conditional Probability

2.1 Joint Events

- Joint event - $A \cap B$ ($A \cap B \neq \emptyset$)
- Joint probability - $P[A \cap B]$
- Marginal probability - $P[B]$

2.2 Conditional Probability

Probability of A when probability B given

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \tag{1}$$

$P[B] \neq 0$

Statistically Independent - P[A] is the same weather or not we know that B has occurred, then A said to be *statistically independent* of the event B.

2.3 Axioms in Conditional Probability

All axioms are in ordinary probability are true in conditional probability

- **Axiom 1**

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \geq 0$$

- **Axiom 2**

$$P[\mathcal{S}] = \frac{P[S \cap B]}{P[B]} = \frac{P[B]}{P[B]} = 1$$

- **Axiom 3**

$$P[A \cup C|B] = P[A|B] + P[C|B]$$

$(A \text{ and } C \text{ are mutually exclusive})$

proof:

$$\begin{aligned} P[A \cup C|B] =& \frac{P[(A \cup C) \cap B]}{P[B]} \\ =& \frac{P[(A \cap B) \cup (C \cap B)]}{P[B]} \\ =& \frac{P[A \cap B]}{P[B]} + \frac{P[C \cap B]}{P[B]} \\ =& P[A|B] + P[C|B] \end{aligned}$$

2.4 The law of total probability

$$P[A] = \sum_{i}^{N} P[A|B_i]P[B_i]$$

$\mathcal{S} = \bigcup_i^N B_i$ $(B_i \cap B_j = \emptyset,\, i \neq j)$ (B_i is a partition of $\mathcal{S}$)

proof:

$$\begin{aligned} P[A] =& P[A \cap \mathcal{S}] \\ =& P\left[A \cap \left(\bigcup_{i}^{N} B_i\right)\right] \\ =& P\left[(A \cap B_1) \cup (A \cap B_2) \cdots \cup (A \cap B_N)\right] \\ =& P\left[(A \cap B_1)\right] + P\left[(A \cap B_2)\right] \ldots P\left[(A \cap B_N)\right] \\ =& P\left[(A|B_1)\right] P\left[B_1\right] + P\left[(A|B_2)\right] P\left[B_2\right] \ldots P\left[(A|B_N)\right] P\left[B_N\right] \\ =& \sum_{i}^{N} P\left[A|B_i\right] P\left[B_i\right] \end{aligned}$$

2.5 Statistically independent events

A and B are statistically independent, then

$$P[A|B] = P[A]$$
$$P[A \cap B] = P[A]P[B]$$

In general events $E_1, E_2 \dots E_N$ define to be statistically independent if

$$P[E_1 E_2 \dots E_N] = P[E_1]P[E_2] \dots P[E_N]$$

2.6 Mutually exclusive events

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = 0$$

2.7 Probability chain rule

Events A, B, C are statistically independent

$$P[A \cap B \cap C] = P[A|(B \cap C)]P[B|C]P[C]$$

2.8 Bayeś Theorem

$$P[B|A] = \frac{P[A|B]P[B]}{P[A]}$$

$P[B|A]$ - Posterior Probability
$P[B]$ - Prior Probability
$P[A|B]$ - Conditional Probability

2.9 Bernoulli Sequence

Bernoulli Trial - Single tossing of a coin with probability p of heads is an example for Bernoulli trial
Bernoulli Sequence - Consecutive independent Bernoulli trials is the Bernoulli sequence

2.10 Binomial Probability Law

$$P[k] = \binom{M}{k} p^k (1-p)^{M-k}$$

M - trials
k - successes
p - probability of successes
P[k] - probability of k successes in M trials

2.11 Geometric Probability Law

$$P[k] = (1-p)^{k-1} p$$

P[k] - k^{th} appearance of first successes
p - probability of successes

2.12 Multinormal Probability Law

Sequence with independent trials with more then 2 outcomes (coin toss as only 2 outcomes, H and T)
$\mathcal{S} = \{1, 2, 3\}$ - given trial can have a 1, 2 or 3. (for the coin with this will H and T only)
p_1, p_2, p_3 - probabilities of getting 1,2 or 3 for a trial ($p_1 + p_2 + p_3 = 1$)

M - Trials
Expecting two 1s ($2p_1$), three 2s ($3p_2$) and one 3s ($1p_3$)
The probability of getting this combinations within M - trials

$$\binom{M}{2, 3, 1} p_1^2 p_2^3 p_3^1$$

$\binom{M}{2,3,1} = \frac{M!}{2!3!1!}$ - Multinomial coefficient

3 Lecture 3: (Chapter 5) Discrete Random Variable

3.1 Discrete Random Variable

Random Variable is a function that maps the sample space $\mathcal{S}$ into a subset of the real line.

A *discrete* random variable is an rv whose possible values either constitute a finite set or else be listed in an infinite sequence in which there is a first element, a second element, and so on.

3.2 Probability Mass Function (pmf)

probability distribution or *probability mass function* of a discrete random variable is define for every number x by $p(x)$

$$p(x) = P(X = x) = P(\forall \in \mathcal{S} : X(s) = x)$$

Conditions

- $0 \leq p(x) \leq 1$
- $\sum_{i=1}^{\infty} p(x_i) = 1$

3.3 Poisson Probability Distribution

A random variable X is said to have a *Poisson Distribution* with parameter $\lambda(\lambda \geq 0)$ if the pmf of X is

$$p(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

λ is frequently a rate per unit time or per unit area

3.4 Approximation of Binomial Probability Mass Function by Poisson pmf

When $M \to \infty$ and $p \to 0$, Binomial pmf become a Poisson pmf ($\lambda = Mp$)

3.5 Transformation of Discrete Random Variable

TODO

3.6 Cumulative Distribution Function (CDF) (*Distribution function*)

$$F_X(x) = P[X \leq x]$$

Note: $P[X \leq x] = P[x_1] + P[x_2] + P[x_3] \cdots + P[x]$

$$P[a < X < b] = F_X(b^-) - F_X(a^+)$$

$$P[a \leq X < b] = F_X(b^-) - F_X(a^-)$$

$$P[a < X \leq b] = F_X(b^+) - F_X(a^+)$$

Note: a^+- *denotes value just slightly larger than a, and* a^- - *denotes value just smaller than b*

Properties of CDF

- $0 \leq F_X(x) \leq 1$
- $\lim_{x \to -\infty} F_X = 0$
- $\lim_{x \to \infty} F_X = 1$
- if $(x_1 < x_2)$ then $F_X(x_1) \leq F_X(x_2)$ - *CDF is monotonically increasing*
- $P(x_1 < X x_2) = F_X(x_2) - F_X(x_1)$ - $(x_1 < x_2)$
- $F_X(x^+) = F_X$ where $x_+ = x + \epsilon$ with $\epsilon \to 0, F_X(x)$ is continuous from the right.

4 Expected Values for Discrete Random Variables ($E[X] = \mu_x$)

4.1 Expected value of a random variable E[X]

Expected value of a random variable is the average value of the outcomes of a large number of experiments.

Average value converge to *expected value*
Expected value of X= expectation of X = average of X = mean of X

$$E[X] = \sum_i x_i p_X(x_i)$$

Expected value = best prediction of the outcome of random experiment of a single trial

Expected value analogues to the center of mass of a system of linearly arranged masses.

4.1.1 Bernoulli Distribution

It is a discrete distribution pertains to an experiment which has exactly two outcomes, which may be labeled {0,1} or {True,False}, for example. One of the outcomes, usually True or 1, has probability p and the other has probability 1-p.

Expected (mean) value E[X] = p
Variance var[X] = p(1-p)

proof:

$$\begin{aligned} E[X] =& \sum_{k=0}^{1} x_i p_X[k] \\ =& 0.(1-p) + 1.p \\ =& p \end{aligned}$$

4.1.2 Binomial Distribution

Consider a random walk in which a unit step to the East occurs with probability p and to the West with probability q = 1 - p. After n steps the probability of taking k steps to the East and n-k to the West is (def)

Expected value $E[X] = Mp$
Variance var(X)=Mp(1-p)

proof:

$$\begin{aligned}
E[X] =& \sum_{k=0}^{M} k p_X[k] \\
=& \sum_{k=0}^{M} k \binom{M}{k} p^k (1-p)^{M-k} \\
=& \sum_{k=0}^{M} k \frac{M!}{(M-k)!k!} p^k (1-p)^{M-k} \\
=& \sum_{k=1}^{M} \frac{M!}{(M-k)!(k-1)!} p^k (1-p)^{M-K} \\
=& Mp \sum_{k'=0}^{M'} \frac{M'!}{(M'-k')!(k'!)} p^{k'} (1-p)^{M'-k'} \\
=& Mp \sum_{k'=0}^{M'} \binom{M'}{k'} p^{k'} (1-p)^{M'-k'} \\
=& Mp
\end{aligned}$$

4.1.3 Poisson Distribution

(def)
Expected value $E[X] = \lambda$
Variance $= \lambda$
Standard Diviation $= \sqrt{\lambda}$

4.1.4 Geometric Distribution

(def:) Expected value $E[X] = \frac{1}{p}$

4.1.5 Gaussian (Normal) Distribution

$$f_X(x; \mu.\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\mu = E[X]$ and $\sigma^2 = \text{var}[X]$

CDF is related to *error function*

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

4.1.6 Rayleigh Distribution

TODO

4.1.7 Conditional Distributions

TODO

4.2 Not all the PMF have expected values

* Discrete random variables with a finite number of values always have E[X]
* Countably infinite numbers of values, a discrete random variable **may not have an E[X]**

Some properties of E[X]

- E[X] is locate at the "center" of the PMF if PMF symmetric about the same point
- IT does NOT generally indicate the most probable value of the random variable
- More than one PMF may have the same E[X] value

4.3 Moments of PMF

- First Moment - Expected value E[X]
- Second Moment - $E[X^2]$
- Third Moment -

4.4 Expected (E[X]) Value for a Function of a Random Variable

Function $Y = g(X)$
Expectation value $E[Y] = \sum_i y_i p_Y[y_i]$

Using function g(x):

$$E[g(x)] = \sum_i g(x) p_Y(y_i)$$

Expectation operator E is linear

$$E[a_1 g_1(X) + a_2 g_2(X)] = a_1 E[g_1(X)] + a_2 E[g_2(X)]$$

Expectation operator does not commute

$$E[g(X)] \neq g(E[X])$$

4.5 Mean Squire Error - MSE

$$E[(X - b)^2]$$

X - true out come of an experiment
b - prediction

4.6 Variance

The variance of a random variable is the expected value of the squared difference from the mean.

$$var(X) = E[(X - E[X])^2]$$
$$var(X) = E[X^2] - E[X]^2$$

4.6.1 Some properties

$$var(c) = 0$$

$$var(X + c) = var(X)$$

$$var(cX) = c^2 var(X)$$

Variance operator is a nonlinear operator

$$var(g_1(x) + g_2(x)) \neq var(g_1(x)) + var(g_2(x))$$

4.7 Characteristic Function

Characteristic Function use to calculate moments of random variables
Characteristic Function of a random variable X

$$\phi_X(\omega) = E[e^{i\omega X}] = \sum_{k=1}^{\infty} p_X[k] e^{i\omega k}$$

$$e^{i\omega X} = \cos(\omega X) + i\sin(\omega X)$$

$$\begin{aligned}
\phi_X(\omega) =& E[e^{i\omega X}] \\
=& E[\cos(\omega X)] + iE[\sin(\omega X)] \\
=& \sum_i p_X[x_i] cos(\omega X) + \sum_i p_X[x_i] sin(\omega X) \\
=& \sum_i p_X[x_i] e^{i\omega X}
\end{aligned}$$

4.7.1 Expected value using Characteristic function

$$\frac{d\phi_X(\omega)}{d\omega} = \sum_{k=-\infty}^{+\infty} p_X[x_i](ik)e^{i\omega k} \tag{2}$$

$$E[X] = \frac{1}{i}\frac{d\phi_X(\omega)}{d\omega}|_{\omega=0} = \sum_{k=-\infty}^{+\infty} p_X[x_i](k)e^{i\omega k}$$

4.7.2 n^{th} moment of a random variable using Characteristic function

$$E[X^n] = \frac{1}{i^n}\frac{d^n\phi_X(\omega)}{d\omega^n}|_{\omega=0}$$

4.8 Some Properties of Characteristic function

- $\phi_X(\omega)$ always exists since

$$|\phi_X(\omega)| < \infty$$

- $\phi_X(\omega)$ is periodic with period 2π

$$\phi_X(\omega + 2\pi m) = \phi_X(\omega)$$

- The PMF may be can recover from the characteristic function.

$$p_X[k] = \int_{-\pi}^{\pi} \phi_X(\omega)e^{-i\omega k}\frac{d\omega}{2\pi}(-\infty < k < \infty)$$

- Convergence of characteristic function guarantees convergence of PMF

4.9 Estimating Expected value and the Variance

$$E[X] = \sum_{k=1}^{N} k p_X[k] = \sum_{k=1}^{N} k(\frac{N_k}{N}) = \frac{1}{N}\sum_{k=1}^{N} kN_k$$

$$var(X) = \frac{1}{N}\sum_{k=1}^{N} x_i^2 - (\frac{1}{N}\sum_{k=1}^{N} x_i)^2$$

5 Functions of a Random Variable

5.1 Jointly Distributed Random Variables

Two random variables that are defined on the *same sample space* S are said to be *jointly distributed*

5.2 Join PMF

$$p_{X,Y}[x_i, y_j] = P[X(s) = x_i, Y(s) = y_j]$$

Properties

- Range of values of joint PMF

$$0 \leq p_{X,Y}[x_i, y_j] \leq 1$$

- Sum of values of joint PMF

$$\sum_{i}^{N_X} \sum_{j}^{N_Y} p_{X,Y}[x_i, y_j] = 1$$

5.3 Marginal PMF

If the join PMF known, then PMF for X and Y ($p_X[x_i], p_Y[y_i]$) can be determine

$$p_X[x_i] = \sum_{j=1}^{\infty} p_{X,Y}[x_k, y_j]$$

$$p_Y[y_i] = \sum_{i=1}^{\infty} p_{X,Y}[x_i, y_k]$$

Join PMF $\Rightarrow$ marginal PMSs

Marginal PMFs $\not\Rightarrow$ marginal PMSs

5.4 Join CDF

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

6 Two Random Variables

6.1 Normal Random Variables

suppose that x and y are normal random variables. The vector $z = [x, y]^T$ is a random point in the plane that can be described by a *mean vector* and *covariance matrix*

6.1.1 mean vector

$$\mu = E[z] = \begin{bmatrix} E[x] \\ E[z] \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$$

6.1.2 covariance martix

$$\Sigma = \begin{bmatrix} E[(x-\mu_x)^2] & E[(x-\mu_x)(y-\mu_y)] \\ E[(x-\mu_x)(y-\mu_y)] & E[(y-\mu_y)^2] \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_x^2 & r\sigma_x\sigma_y \\ r\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$$

6.2 Conditional Distribution

$$F_z(z|M) = \frac{P(z \leq z|M)}{P(M)}$$

7 Modeling of Photone Detectors

7.1 Contrast

Irradiance - power of electromagnetic radiation per unit area at a surface.

Brightness of a Object is

$$C = \frac{|\bar{q}_b - \bar{q}_o|}{\bar{q}_b}$$

$\bar{q}_o$ - Irradiance of the object
$\bar{q}_b$ - Irradiance of the background

7.2 Rose Model

Signal - (S_{Rose}) $= |\bar{q}_b - \bar{q}_o|A$
Noise - $\sigma_b = \sqrt{A\bar{q}_b}$

7.2.1 Signal to Noice ratio

$$SNR_{Rose} = \frac{|\bar{q}_b - \bar{q}_o|A}{\sqrt{A\bar{q}_b}} = C\sqrt{A\bar{q}_b}$$

7.3 Detective Quantum Efficiency

DQE - Ability to detect a random signal against the background of random radiation.

$$DQE = \frac{SNR^2_{ouput}}{SNR^2_{input}} = \frac{(\Delta X)^2\sigma_Y^2}{(\Delta Y)^2\sigma_X^2} = \frac{G\sigma_X^2}{\sigma_Y^2}$$

7.4 Defining Detector Performance

- Statistical
- SNR
- Minimum integration time

7.4.1 Statistical

DEQ is define as the ratio of the variance of the tw estimates.

$$DEQ = \frac{variance - of - the - ideal - estimator}{variance - of - actual - estimator} = \frac{\sigma^2_{q-ideal}}{\sigma^2_{q-actual}}$$

Maximum DQE = 1

7.4.2 SNR

$$SNR = \frac{\bar{X}}{\sigma_X}$$

$\bar{X} = E[X]$