LECTURE 7

Facts From Probability Theory

Notation:

$P(A)$ = Probability that the event described by A occurs.
$P(A|B)$ = Probability that A occurs given that B has occurred (conditional probability);

$A \cap B$ = *A and B*.
$A \cup B$ = *A or B*.

Note:

(a) $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$.

(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, always.

(c) If A and B are mutually exclusive, i.e. $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.

(d) If $A_1 \cap A_2 = \emptyset$ (mutually exclusive A and B) then $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$.

(e) If $A \subset B$, i.e. A implies B, then $P(A|B) = \dfrac{P(A)}{P(B)}$.

(f) $P(A \cap B) = P(A)P(B)$ **if and only if** the events A and B are independent.

Cumulative *distribution function* (c.d.f.): $F_X(x) = P(X \le x)$

Two types of random variables:- discrete-type and continuous-type

Discrete-type X:

$F_X(x)$ is piece-wise constant, i.e. X takes on values from a discrete set $\{x_1, x_2, \ldots\}$ and $F_X(x) = \sum_{x_k \le x} P(X = x_k)$.

Continuous-type X:

there exists a continuous function $f_X(x)$, called *probability density function* (p.d.f.) for X, such that

(g) $P(X \le x) = F_X(x) = \int_{-\infty}^{x} f_X(u) du$.

The c.d.f. $F_X(x)$ and p.d.f. $f_X(x)$ for a continuous r.v. X are related through $\frac{d}{dx} F_X(x) = f_X(x)$ which holds at every x where the derivative exists.

Note:

(h)
$$\begin{aligned} P(X > x) &= 1 - F_X(x);, && \text{[always, follows from } P(X > x) + P(X \le x) = 1\text{]} \\ &= \sum_{x_k > x} P(X = x_k) && \text{[if } X \text{ is a discrete-type r.v.]} \\ &= \int_{x}^{+\infty} f_X(u) du. && \text{[if } X \text{ is a continuous-type r.v.]} \end{aligned}$$

(i)
$$\begin{aligned} P(x < X \le y) &= F_X(y) - F_X(x); && \text{[always]} \\ &= \sum_{x < x_k \le y} P(X = x_k) && \text{[if } X \text{ is a discrete-type r.v.]} \\ &= \int_{x}^{y} f_X(u) du. && \text{[if } X \text{ is a continuous-type r.v.]} \end{aligned}$$

In particular, for continuous X:

$$\begin{aligned} P(x \le X \le x + \Delta x) &= F_X(x + \Delta x) - F_X(x) \\ &\simeq \frac{d}{dx} F_X(x) \times \Delta x, \quad \text{for small } \Delta x, \\ &= f_X(x) \Delta x \end{aligned}$$

(j) If X is a continuous-type random variable then

$$P(x_1 < X \le x_2) = P(x_1 \le X \le x_2) = P(x_1 \le X < x_2) = P(x_1 < X < x_2).$$

Expectation (mean value) of X:

$$\begin{aligned} E(X) &= \sum_{x_k} x_k P(X = x_k), && \text{if } X \text{ is a discrete-type random variable;} \\ &= \int_{-\infty}^{+\infty} u \, f_X(u) \, du, && \text{if } X \text{ is a continuous-type random variable.} \end{aligned}$$

Note: $E(X + Y) = E(X) + E(Y)$ and $E(\alpha X) = \alpha E(X)$.

Variance of X (the mean quadratic deviation form the mean value) :

$$\mathrm{var}(X) = E([X - E(X)]^2) = E(X^2) - [E(X)]^2$$

Note: $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\,\mathrm{cov}(X, Y)$ and $\mathrm{var}(\alpha X) = \alpha^2 \, \mathrm{var}(X)$,

where $\mathrm{cov}(X, Y) = E([X - E(X)][Y - E(Y)])$ is the covariance of X and Y. If X and Y are independent $\mathrm{cov}(X, Y) = 0$ and $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y)$.

Note: (Tchebyshev's inequality, explains the meaning of $E(X)$ and $\mathrm{var}(X)$)

$$P(\, |X - E(X)| \ge \varepsilon \,) \le \frac{\mathrm{var}(X)}{\varepsilon^2}$$

Conditioning a random variable X by the event $X > t$.

If X is a continuous-type random variable and t belongs to the range of its values, we can define a new random variable $(X-t)|(X>t)$ (reads $X-t$ given $X>t$) whose probability distribution is the distribution of $X-t$ conditioned by the event $X>t$, i.e.

$$\begin{aligned} F_{(X-t)|(X>t)}(s) &\stackrel{\text{def.}}{=} P(X-t \le s | X > t) \\ &= P(t < X \le t+s | X > t) \end{aligned}$$

This new random variable is obviously non-negative and its p.d.f. $f_{(X-t)|(X>t)}$ is obtained by calculating $P(s < X-t \le s+\Delta s \mid X>t)$ for small Δs and positive s:

$$\begin{aligned} f_{(X-t)|(X>t)}(s) \times \Delta s &\simeq P(s < (X-t) \le s+\Delta s \mid X>t) && \text{[by (i)]} \\ &= \frac{P(s < X-t \le s+\Delta s)}{P(X>t)} && \text{[by (a) and (e)]} \\ &= \frac{P(s+t < X \le s+t+\Delta s)}{P(X>t)} \\ &\simeq \frac{f_X(s+t)\times\Delta s}{P(X>t)}, && \text{[by (i)]} \end{aligned}$$

hence

(k) $$f_{(X-t)|(X>t)}(s) = \frac{f_X(s+t)}{P(X>t)} = \frac{f_X(s+t)}{1-F_X(t)} \text{ if } s \ge 0 \text{ and } f_{(X-t)|(X>t)}(s) = 0 \text{ if } s<0.$$

The conditional expectation of $X-t$ given $X>t$ is denoted by $E(X-t|X>t)$ and

(l) $$E(X-t|X>t) = \int_0^\infty s\, f_{(X-t)|(X>t)}(s)\,ds = \frac{\int_0^\infty s\, f_X(s+t)\,ds}{P(X>t)} = \frac{\int_t^\infty (u-t)\, f_X(u)\,du}{P(X>t)}.$$

Convention:

We will write $X \sim$ (distribution) to express the fact that X is a random variable with the specified distribution.

Examples of distributions:

1. Bernoulli distribution (discrete-type): $X \sim \text{Bernoulli}(p)$,
   X takes on either 1 or 0 (success or failure); $P(X=1)=p$, $P(X=0)=q$; $p+q=1$;
   $E(X)=p$, $\text{var}(X)=pq$

2. Binomial distribution (discrete-type): $X \sim \text{Bin}(n,p)$,
   X takes on the values $0,1,2,\dots n$; $P(X=k) = C_k^n p^k q^{n-k}$, $k=0,1,\dots,n$; $p+q=1$;
   $E(X)=np$, $\text{var}(X)=npq$

   If $X_1, X_2, \dots X_n$ are mutually independent and $X_j \sim \text{Bernoulli}(p)$ for all j then

$$X_1+X_2+\dots X_n \sim \text{Bin}(n,p).$$

3. Uniform distribution on $[0,1]$ (continuous-type): $X \sim \text{Uniform}[0,1]$,
   X can take on any value between 0 and 1; $P(a \le X \le b) = b-a$ for any $0 \le a < b \le 1$.
   p.d.f.: $f_X(x) = 1$ if $x \in [0,1]$ and $f_X(x)=0$ otherwise.
   $E(X)=\frac{1}{2}$, $\text{var}(X)=\frac{1}{3}$

4. Exponential distribution (continuous-type): $X \sim \text{Exp}(\lambda)$,
   X can take on any non-negative value;
   p.d.f.: $f_X(x) = \lambda e^{-\lambda x}$ if $x \ge 0$ and $f_X(x)=0$ otherwise
   c.d.f.: $F_X(x) = 1-e^{-\lambda x}$ if $x \ge 0$ and $F_X(x)=0$ otherwise
   $E(X)=\frac{1}{\lambda}$, $\text{var}(X)=\frac{1}{\lambda^2}$

   If $X \sim \text{Exp}(\lambda)$ then $(X-t)|(X>t) \sim \text{Exp}(\lambda)$ as well:

$$\begin{aligned} f_{(X-t)|(X>t)}(s) &= \frac{f_X(s+t)}{1-F_X(t)} = \frac{\lambda e^{-\lambda(s+t)}}{e^{-\lambda t}} \\ &= \lambda e^{-\lambda s} = f_X(s) && s,t \ge 0. \end{aligned}$$

   Also, if $X \sim \text{Exp}(\lambda)$ then $P(X-t>s|X>t) = P(X>s)$.