Lectures 22-23

Theorem 0.1 *If the random variable X has finite moments upto order n. Then Φ has continuous derivatives upto order n.More over*

$$i^k EX^k \;=\; \Phi_X^{(k)}(0),\;\; k = 1, 2, \dots, n\,.$$

Proof.
Consider

$$\frac{\Phi_X(t+h) - \Phi_X(t)}{h} \;=\; E\Big[e^{itX}\frac{(e^{ihX}-1)}{h}\Big]$$

since $|e^{ihx} - 1| \leq 2|hx|$, we get

$$\Big|e^{itX}\frac{(e^{ihX}-1)}{h}\Big| \;\leq\; 2|X|$$

and $E|X| < \infty$. Hence by Dominated Convergence theorem

$$\lim_{h\to 0} E\Big[e^{itX}\frac{(e^{ihX}-1)}{h}\Big] \;=\; E[iXe^{itX}]\,.$$

Therefore

$$\Phi_X'(t) \;=\; E[iXe^{itX}]\,.$$

Put $t = 0$, we get

$$\Phi_X^{(1)}(0) \;=\; i\,EX\,.$$

For higher order derivatives, repeat the above arguments. □

Theorem 0.2 *(Inversion theorem) Let X be a random variable with distribution function F and characteristic function $\phi_X(\cdot)$. Then*

$$F(b) - F(a) \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-ita} - e^{-itb}}{it}\Phi_X(t)dt,$$

whenever a, b are points of continuity of F.

Proof. (Reading exercise) *Before proceeding towards the sketch of proof, a word about the integral on the rhs. The integral on the rhs is interpreted as an improper Riemann integral which may not be (in general) absolutely integrable. At this stage, student may not worry about this.*

Consider

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-ita}-e^{-itb}}{it}\Phi_X(t)dt &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-ita}-e^{-itb}}{it}Ee^{itX}dt \\
&= \frac{1}{2\pi}E\int_{-\infty}^{\infty}\frac{e^{-ita}-e^{-itb}}{it}e^{itX}dt \qquad (0.1)\\
&= E\int_{-\infty}^{\infty}\frac{e^{it(X-a)}-e^{it(X-b)}}{2\pi it}dt\,.
\end{aligned}$$

The second equality follows from the change of order of integration (This in fact, requires to consider the integrals on finite intervels say for example $[-T,T]$(i.e. proper integrals) and use change of variable formula there and then let $T\to\infty$). Now

$$\int_{-\infty}^{0}\frac{e^{it(X-a)}-e^{it(X-b)}}{2\pi it}dt = \int_{0}^{\infty}\frac{e^{-it(X-a)}-e^{-it(X-b)}}{2\pi it}dt \qquad (0.2)$$

Hence, using $2i\sin\theta = e^{i\theta}-e^{-i\theta}$, we get

$$2i(\sin t(X-a)-\sin t(X-b)) = e^{it(X-a)}-e^{it(X-b)}-(e^{-it(X-a)}-e^{-it(X-b)}). \qquad (0.3)$$

Combining (0.2) and (0.3), we get

$$\int_{-\infty}^{\infty}\frac{e^{it(X-a)}-e^{it(X-b)}}{2\pi it}dt = \frac{1}{\pi}\int_0^{\infty}\frac{\sin\, t(X-a)}{t}dt - \frac{1}{\pi}\int_0^{\infty}\frac{\sin\, t(X-b)}{t}dt\,. \qquad (0.4)$$

Using

$$\int_0^{\infty}\frac{\sin\,\alpha x}{x}dx = \frac{\pi}{2}sgn(\alpha)$$

we get

$$\frac{1}{\pi}\int_0^{\infty}\frac{\sin\, t(X-a)}{t}dt = \begin{cases} \frac{1}{2} & \text{if} \quad X>a \\ 0 & \text{if} \quad X=a \\ -\frac{1}{2} & \text{if} \quad X<a\,, \end{cases} \qquad (0.5)$$

where

$$sgn(\alpha) = \begin{cases} -1 & \text{if} \quad \alpha<0 \\ 0 & \text{if} \quad \alpha=0 \\ 1 & \text{if} \quad \alpha>0\,. \end{cases}$$

Similarly, the other integral. Combining (0.1), (0.4) and (0.5), we complete the proof.

Remark 0.1 *In this remark, I will give the computation of the integral* $\int_0^\infty \frac{\sin \alpha x}{x} dx$*. It is enough to compute the Dirichlet integral* $\int_0^\infty \frac{\sin x}{x} dx$ *because for other values of* α*, the integral follows from the Dirichlet integral easily. For example,*

$$\int_0^\infty \frac{\sin(-x)}{x} dx \ = \ -\int_0^\infty \frac{\sin x}{x} dx.$$

Before even proceeding to compute this, I will show that $\int_0^\infty \frac{\sin x}{x} dx$ *is not absolutely integrable but is integrable. First recall that*

$$\int_0^\infty \frac{\sin x}{x} dx := \lim_{\varepsilon \to 0, T \to \infty} \int_\varepsilon^T \frac{\sin x}{x} dx.$$

Consider

$$\begin{aligned} \int_1^T \frac{\sin x}{x} dx \ &= \ \int_1^T \frac{1}{x}(1 - \cos x)' dx \\ \text{integration by parts} \ &= \ \Big[\frac{1 - \cos x}{x}\Big]_1^T + \int_1^T \frac{1 - \cos x}{x^2} dx \\ &= \ \frac{1 - \cos T}{T} - 1 + \cos 1 + \int_1^T \frac{1 - \cos x}{x^2} dx. \end{aligned}$$

Hence

$$\int_1^\infty \frac{\sin x}{x} dx \ = \ \lim_{T \to \infty} \int_1^T \frac{\sin x}{x} dx \ = \ \cos 1 - 1 + \int_1^\infty \frac{1 - \cos x}{x^2} dx$$

exists. Now since

$$\lim_{x \downarrow 0} \frac{\sin x}{x} \ = \ 1$$

the function $f(x) = \frac{\sin x}{x}, x > 0, \ = 1, x = 0$ *is Riemann integrable and hence*

$$\lim_{\varepsilon \to 0} \int_\varepsilon^1 \frac{\sin x}{x} dx \text{ exists.}$$

Combining above arguments it follows that

$$\lim_{\varepsilon \to 0, T \to \infty} \int_\varepsilon^T \frac{\sin x}{x} dx$$

exists. i.e. Dirichlet integral exists as improper Riemann integral. Now we will see that it is not absolutely integrable. Note that

$$\begin{aligned}\int_0^\infty \frac{|\sin x|}{|x|}dx &\geq \int_0^\infty \frac{|\sin^2 x|}{|x|}dx \\ &= \int_0^\infty \frac{|1-\cos 2x|}{2x}dx \\ &= \int_0^\infty \frac{1-\cos 2x}{2x}dx \\ &\geq \int_1^\infty \frac{1-\cos 2x}{2x}dx.\end{aligned}$$

The last integral has two parts in which first one is known to diverge to ∞ and the second one converges using the same arguement as for the Dirichlet integral. Hence the integral on the lhs also diverges.

Now we compute the Dirichlet integral. To this end, first we need the value of the following. Consider for any given $u > 0$

$$\begin{aligned}\int_0^\infty e^{ix}e^{-xu}dx &= \lim_{T\to\infty}\int_0^T e^{(i-u)x}dx \\ &= \lim_{T\to\infty}\left[\frac{e^{(i-u)x}}{i-u}\right]_0^T \\ &= \frac{-1}{i-u} = \frac{u+i}{1+u^2}.\end{aligned}$$

Hence equating the real and imaginary parts, we get the values of the following improper Riemann integrals

$$\int_0^\infty \cos xe^{-xu}dx = \frac{u}{1+u^2}, \int_0^\infty \sin xe^{-xu}dx = \frac{1}{1+u^2}.$$

Now consider

$$\begin{aligned}\int_0^\infty \frac{\sin x}{x}dx &= \int_0^\infty \sin x\int_0^\infty e^{-xu}dudx \\ &= \int_0^\infty\int_0^\infty \sin xe^{-xu}dxdu \\ &= \int_0^\infty \frac{1}{1+u^2}du = \frac{\pi}{2}.\end{aligned}$$

Now we see a special case of the inversion formula, i.e. when X has a pdf f.

Corollary 0.1 *Let X be a continuous random variable with a pdf f, then*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\Phi_X(t)dt, \ x \in \mathbb{R}.$$

Proof: (Reading exercise) I will give an idea of the proof. The proof is all about computing the derivative of the distribution function F using the inversion formula. i.e.,

$$\begin{aligned}
\lim_{h\to\infty} \frac{F(x+h)-F(x)}{h} &= \lim_{h\to\infty} \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-itx}-e^{-it(x+h)}}{ith}\Phi_X(t)dt \\
&= \lim_{h\to\infty} \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{(1-e^{-ith})}{ith}\Phi_X(t)dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \lim_{h\to\infty} e^{-itx}\frac{(1-e^{-ith})}{ith}\Phi_X(t)dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\Phi_X(t)dt.
\end{aligned}$$

Note in the above, the first equality is from the inversion formula and the third equality is by an indirect application of dominated convergence theorem, this is in fact the delicate part. [Here one need to first look at integrals from $t = -T$ and $t = T$ and apply DCT and then let $T \to \infty$ in view of the fact that the integrals exists as improper Riemann integrals.]

Theorem 0.3 *(Uniqueness Theorem)*

Let X_1, X_2 be two random variables such that $\Phi_{X1} \equiv \Phi_{X2}$. Then X_1, X_2 have same distribution.

Proof:

Using Inversion theorem, we have

$$F_1(b) - F_1(a) = F_2(b) - F_2(a)$$

for all $a, b \in \mathbb{R}$ such that F_1, F_2 are continuous at a and b.

Now let $a \to -\infty$, we have

$$F_1(b) = F_2(b)$$

for all b at which F_1 and F_2 are continuous.

Therefore

$$F_1(x) = F_2(x), \ x \in \mathbb{R} \text{ (Exercise)}$$

Remark 0.2 *Theorem 0.3 points out the following important fact. Characteristic function is about the distribution of the random variable not about the random variable itself. Hence it is more appropriate to tell 'characteristic function of a distribution (i.e. a probability mesaure on* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *rather than 'characteristic function of random variable'*

Now we will state a result which is useful, its proof is beyond the scope of this course.

Theorem 0.4 *(Continuity Theorem) Let* X_n, X *be random variables on* $(\Omega, \mathcal{F}, P)$ *such that,*

$$\lim_{n\to\infty} \Phi_{X_n}(t) \ = \ \Phi_X(t), t \in \mathbb{R}.$$

Then $F_{X_n}(x) \to F_X(x)$ *for all* $x \in \mathbb{R}$ *such that* F *is continuous at* x*.*

0.0.1 Some fundamental Inequalities

We end this chapter with some fundamental inequalities. We begin with **Cauchy-Schwarz inequality**.

Lemma 0.1 *Let* X *and* Y *be random variables with finite second moment. Then* XY *has finite mean and satisfies*

$$E|XY| \ \le [EX^2]^{\frac{1}{2}} \ [EY^2]^{\frac{1}{2}}.$$

Proof: If $EX^2 = 0$ or $EY^2 = 0$, then proof is easy (exercise [1]). So we assume that $EX^2, EY^2 \neq 0$. Consider for $\lambda = \frac{E|XY|}{EY^2}$ [2],

$$\begin{aligned} 0 \le E[|X| - \lambda|Y|]^2 &= EX^2 - 2\lambda E|XY| + \lambda^2 EY^2 \\ &= EX^2 - \frac{(E|XY|)^2}{EY^2}. \end{aligned}$$

Hence the inequality follows. $\square$

Lemma 0.2 *(Jensen's inequality) Let* $\varphi : \mathbb{R} \to \mathbb{R}$ *be a convex function and* X *be such that* EX *and* $E\varphi(X)$ *exists. Then*

$$\varphi(EX) \ \le \ E\varphi(X).$$

[1]Hint: When $EX^2 = 0$, use Markov inequality to show that $X^2 = 0$ a.s.

[2]The function $f(\lambda) = EX^2 - 2\lambda E|XY| + \lambda^2 EY^2$ has a unique minimum at $\frac{E|XY|}{EY^2}$

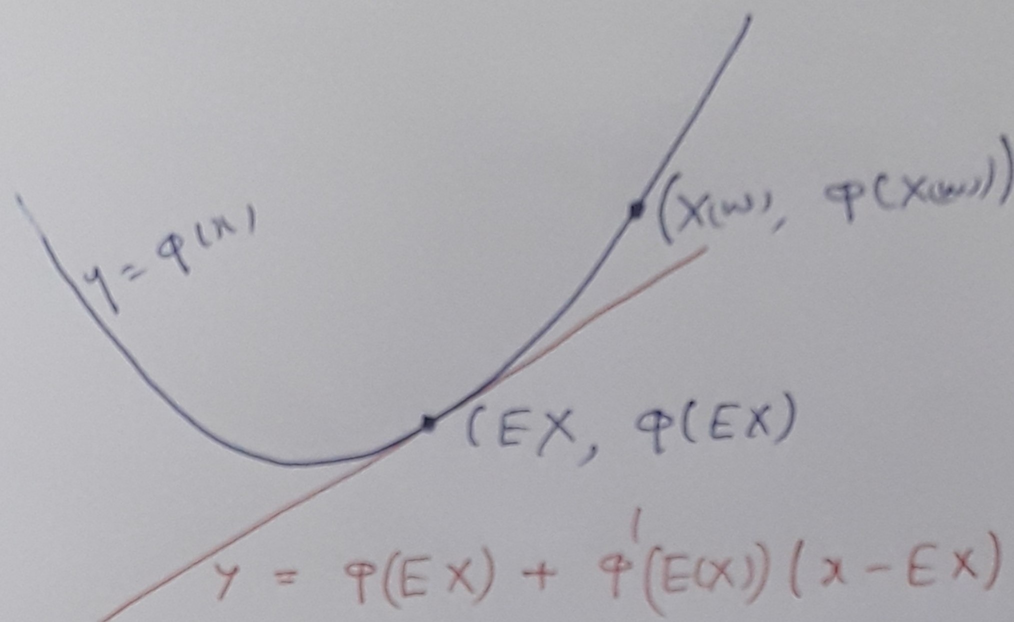
Proof:

$$y = \varphi(EX) + \varphi'(EX)(x - EX)$$

denote the equation of the tangent to $y = \varphi(x)$ at $(EX, \varphi(EX))$. Then for any point $(X, \varphi(X))$ on the graph of $y = \varphi(x)$ lies above the tangent line, see the figure.

Hence

$$\begin{aligned} \varphi(X) &\geq \varphi(EX) + \varphi'(EX)(X - EX) \\ \Rightarrow E\varphi(X) &\geq \varphi(EX) + 0. \end{aligned}$$



y = φ(x)
(X(ω), φ(X(ω)))
(EX, φ(EX)
y = φ(EX) + φ'(E(X))(x − EX)

This completes the proof. $\square$

Given two discrete probability distributions $P = (p_1, p_2, \cdots)$ and $Q = (q_1, q_2, \cdots)$, we define the Kullback-Leibler divergence from P and Q as

$$D(P,Q) = \sum_k p_k \ln(\frac{p_k}{q_k}) = -\sum_k p_k \ln(\frac{q_k}{p_k})$$

(A notion very useful in information theory and is also known as relative entropy) Using Jensen's inequality, we can see that $D(P,Q) \geq 0$.

[Define the random variable X as follows $X = \sum_k \frac{q_k}{p_k} I_{A_k}$ with probability $P(A_k) = p_k$. Then $EX = 1$ and

$$D(P,Q) = E[-\ln(X)] \geq -\ln EX = 0$$

The last inequality follows, since $-\ln x, x > 0$ is convex.]

Power of Markov inequality and Chebyschev's inequality are in their generality not in their sharpness. But one can use Markov inequality to get some nice estimates called the exponential inequalities. Here one need to use the structure of the distribution. Following is an important manifestation of this.

Theorem 0.5 *(Hoeffding's inequality-Binomial case) Let X denote Binomial $(n, \frac{1}{2})$, then*

$$P\Big\{\Big|X - \frac{n}{2}\Big| \geq nt\Big\} \leq 2e^{-2nt^2}, t > 0.$$

Proof: Note $X = X_1 + \cdots + X_n$, where X_i's are i.i.d. Bernoulli $(\frac{1}{2})$.
Fix $\lambda > 0$.

$$\begin{array}{rcl} P\big(X \geq \frac{n}{2} + nt\big) & = & P\Big(\lambda(X - \frac{n}{2} - nt) \geq 0\Big) \\ & = & P\Big(e^{\lambda(X - \frac{n}{2} - nt)} \geq 1\Big) \\ & = & P\Big(e^{\lambda(X - \frac{n}{2})} \geq e^{\lambda nt}\Big) \\ \text{(use Markov inequality)} & \leq & e^{-\lambda nt} E[e^{\lambda(X - \frac{n}{2})}] \\ & = & e^{-\lambda nt} \Pi_{i=1}^n E[e^{\lambda(X_i - \frac{1}{2})}] \\ & = & e^{-\lambda nt}\Big(\dfrac{e^{-\frac{\lambda}{2}} + e^{\frac{\lambda}{2}}}{2}\Big)^n \\ & \leq & e^{-\lambda nt} e^{\frac{n\lambda^2}{8}} \\ & = & e^{-\lambda nt + \frac{\lambda^2 n}{8}}. \end{array}$$

In the last inequality, we use the following.

$$\begin{aligned}\Big(\frac{e^{-\frac{\lambda}{2}}+e^{\frac{\lambda}{2}}}{2}\Big)^2 &= \frac{1}{4}\Big(e^{\frac{\lambda^2}{4}}+2+e^{-\frac{\lambda^2}{4}}\Big)\\ &\leq e^{\frac{\lambda^2}{4}}.\end{aligned}$$

Hence

$$P\Big(X \geq \frac{n}{2}+nt\Big) \;\leq\; \inf_{\lambda>0} e^{-\lambda nt+\frac{\lambda^2 n}{8}} \;=\; e^{-2nt^2}.$$

A similar arguement implies

$$P\Big(X \leq \frac{n}{2}-nt\Big) \;\leq\; e^{-2nt^2}.$$

Note that, the above proof contains a general procedure. We will make this explicit in the following general 'exponential' inequality.

Theorem 0.6 *For any random variable X, we have the following tail bound.*

$$P\big(X \geq t\big) \leq \inf_{\lambda \geq 0} E[e^{\lambda(X-t)}],\ t>0.$$

Proof It follows easily from

$$\begin{aligned}P\big(X \geq t\big) &= P\big(e^{\lambda(X-t)} \geq 1\big)\\ \text{(Markov inequality)} &\leq E[e^{\lambda(X-t)}].\end{aligned}$$

Theorem 0.7 *(Chernoff's inequality) Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables, with mgf $M(\cdot)$. Then*

$$P\big(S_n \geq t\big) \;\leq\; \inf_{\lambda>0}\Big(e^{-\lambda t/n}M(\lambda)\Big)^n,\ P\big(S_n \leq -t\big) \;\leq\; \inf_{\lambda>0}\Big(e^{-\lambda t/n}M(-\lambda)\Big)^n.$$

Proof follows from Theorem 0.6 by taking $X = S_n$ and $X = -S_n$.

Example 0.1 *Consider the case when $S_n \sim$ Binomial (n,p). Note*

$$M(\lambda) = 1-p+pe^{\lambda}.$$

Set

$$f(\lambda) \;=\; e^{-\lambda t/n}(1-p+pe^{\lambda}).$$

Now verify that f attains its minimum at

$$\lambda \;=\; \ln\Big[\frac{\frac{(1-p)t}{n}}{p(1-\frac{t}{n})}\Big]$$

which is positive for $p \leq t/n \leq 1$*. Hence for* $np \leq t \leq n$*.*

$$\begin{aligned}
\inf_{\lambda>0} f(\lambda) &= \Big(\frac{t/n}{p}\Big)^{-\frac{t}{n}}\Big(\frac{1-t/n}{1-p}\Big)^{-(1-t/n)} \\
&= e^{-\frac{t}{n}\ln\frac{t/n}{p}-(1-\frac{t}{n})\ln\frac{1-t/n}{1-p}} \\
&= e^{-D(Q_{t/n},P)}.
\end{aligned}$$

where $P = (p, 1-p), Q_{t/n} = (t/n, 1-t/n)$ *and* $D(Q_{t/n}, P)$ *is the Kullback-Leibler divergence from* Q_n *and* P*. This gives an upper tail bound*

$$P\Big(S_n \geq t\Big) \;\leq\; e^{-nD(Q_{t/n},P)}.$$

Follwing information is useful to further investigate the above inequality. Set

$$I_p(\alpha) = \alpha \ln\frac{\alpha}{p} + (1-\alpha)\ln\frac{1-\alpha}{1-p},\ p \leq \alpha \leq 1.$$

Then $I_p(p) = 0, I_p(1) = \frac{1}{p}$ *and* I_p *is an increasing function (exercise).*

Chapter 8 : Conditional distribution and expectation

The notion of conditional densities are intended to give a quantification of dependence of one random variable over the other if the random variables are not independent. One of the main use of conditional expectation and probabilities is the 'conditioning argument' to compute expectations which involves multiple random variables. Our approach is to define conditional distributions (pmf/pdf), conditional probabilities and finally conditional expectations for various cases such as discrete, continuous and a combination of both. This is an elementary treatment and hence doesn't cover conditional probabilities in its full generality.

0.1 Conditional pmf of discrete random variables

In this case, definition of conditional pmf directly follows from the definition of conditional probabilities of events.

Definition 0.1 *Let X, Y be two discrete random variables with joint pmf f. Then the conditional pmf of Y given X denoted by $f_{Y|X}$ is defined as*

$$f_{Y|X}(y|x) = \begin{cases} P(Y=y|X=x) & \text{if} \quad P(X=x) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

Intuitively, $f_{Y|X}$ means the pmf of Y given the information about X. Here information about X means knowledge about the occurrence (or non occurrence) of $\{X = x\}$ for each x. One can rewrite $f_{X|Y}$ in terms of the pmfs as follows.

$$f_{Y|X}(y|x) = \begin{cases} \dfrac{f(x,y)}{f_X(x)} & \text{if} \quad f_X(x) \neq 0 \\ 0 & \text{otherwise}\,. \end{cases}$$

Definition of $f_{X|Y}$ is similar. Also one can relate $f_{Y|X}$ and $f_{X|Y}$ through the Bayes theorem as follows.

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) f_Y(y)}{f_X(x)},$$

where f_Y denote the marginal distribution of Y.

0.2 Conditional pdf of continuous random variables

When both X and Y are continuous random variables with a joint pdf f. Let f_X, f_Y denote the respective marginal pdfs of X and Y.

First we need a definition for $P\big(Y \leq y|X = x\big)$, since $\dfrac{P(X = x, Y \leq y)}{P(X = x)}$ doesn't make sense. Let us define $P\big(Y \leq y \,|\, X = x\big)$ as follows. For $f_X(x) > 0$, define

$$P\big(Y \leq y \,|\, X = x\big) \;=\; \lim_{h\to 0} P\big(Y \leq y \,|\, x - h < X \leq x + h\big)$$

and equal to 0 otherwise. Now

$$\begin{aligned}
\lim_{h\to 0} P\big(Y \leq y \,|\, x - h < X \leq x + h\big) &= \lim_{h\to 0} \frac{P\big(Y \leq y,\, x - h < X \leq x + h\big)}{P\big(x - h < X \leq x + h\big)} \\
&= \lim_{h\to 0} \frac{F(x+h, y) - F(x-h, y)}{F_X(x+h) - F_X(x-h)} \\
&= \frac{\lim_{h\to 0} \frac{1}{2h}(F(x+h, y) - F(x-h, y))}{\lim_{h\to 0} \frac{1}{2h}(F_X(x+h) - F_X(x-h))} \\
&= \frac{\frac{\partial F(x,y)}{\partial x}}{f_X(x)} \;=\; \frac{1}{f_X(x)} \int_{-\infty}^{y} f(x,t)dt \\
&= \int_{-\infty}^{y} \frac{f(x,t)}{f_X(x)} dt
\end{aligned}$$

Hence, we define the conditional distribution function $F_{Y|X}(y|x)$ as follows,

$$F_{Y|X}(y|x) \;=\; P\big(Y \leq y \,|\, X = x\big) \;=\; \begin{cases} \int_{-\infty}^{y} \frac{f(x,t)}{f_X(x)} dt & \text{if } \; f_X(x) > 0, \\ 0 & \qquad \text{otherwise} \end{cases}$$

A similar computation leads to

$$P\big(X \leq x \,|\, Y = y\big) \;=\; \int_{-\infty}^{x} \frac{f(s,y)}{f_Y(y)} ds$$

when ever $f_Y(y) > 0$. So we define the conditional distribution function $F_{X|Y}(x|y)$ as follows,

$$F_{X|Y}(x|y) \;=\; \begin{cases} P\big(X \leq x \,|\, Y = y\big) \;=\; \int_{-\infty}^{y} \frac{f(s,y)}{f_Y(y)} ds, \text{ if } f_Y(y) > 0, \\ 0 \qquad\qquad\qquad\qquad\quad \text{otherwise} \end{cases}$$

Now we define $f_{Y|X}$ as the density of $F_{Y|X}(y|x)$ which is $\dfrac{f(x,y)}{f_X(x)}$ from the above. Similarly $f_{X|Y}(x|y)$ is defined. i.e.

Definition 0.2 *Let X, Y are continuous random variables with joint pdf f. The conditional pdf of Y given X is defined as*

$$f_{Y|X}(y|x) \;=\; \begin{cases} \dfrac{f(x,y)}{f_X(x)} & \text{if} \quad 0 < f_X(x) < \infty \\ 0 & \text{if} \quad \text{otherwise} \end{cases}$$

and the conditional pdf of X given Y is defined as

$$f_{X|Y}(x|y) \;=\; \begin{cases} \dfrac{f(x,y)}{f_Y(y)} & \text{if} \quad 0 < f_Y(y) < \infty \\ 0 & \text{if} \quad \text{otherwise}\,. \end{cases}$$

This leads to the following Bayes' theorem for densities

$$f_{Y|X}(y|x)f_X(x) \;=\; f_{X|Y}(x|y)f_Y(y). \tag{0.6}$$

0.3 Conditional pmf/pdf: for a combination of random variables

Now we discuss the case when X is discrete and Y is a continuous random variable. Let f_X denote the pmf of X, Y has a pdf f_Y and F denote the joint distribution function of X and Y. Consider the probabilites

$$G(x,y) \;=\; P\big(X = x, Y \le y\big), x \in D_X, y \in \mathbb{R},$$

where D_X is the set of discontinuities of the distribution function F_X of X. Then it is easy to see that

$$F(x,y) \;=\; \sum_{x' \in D_X, x' \le x} G(x', y).$$

Definition 0.3 *The conditional distribution function of Y given $X = x$ denoted by $F_{Y|X}(y|x)$ is defined as*

$$F_{Y|X}(y|x) \;:=\; P\big(Y \le y \,|\, X = x\big) = \frac{G(x,y)}{f_X(x)},\; x \in D, y \in \mathbb{R}.$$

Now we define the conditional pdf of Y given $X = x$ is defined as the function $f_{Y|X}(y|x)$ satisfying

$$F_{Y|X}(y|x) = \int_{-\infty}^{y} f_{Y|X}(z|x)dz, y \in \mathbb{R}.$$

Clearly, if G is differentiable with respect to y for each $x \in D$, then

$$f_{Y|X}(y|x) = \frac{1}{f_X(x)} \frac{dG(x,y)}{dy}.$$

As in the previous section, we define conditional pmf function of X given $Y = y$ denoted by $f_{X|Y}(x|y)$ as follows.

$$F_{X|Y}(x|y) = \lim_{h \to 0} P\big(X = x | y - h < Y \leq y + h\big)$$

when ever $f_Y(y) > 0$ and equal to 0 otherwise. Now

$$\begin{aligned} \lim_{h \to 0} P\big(X = x | y - h < Y \leq y + h\big) &= \lim_{h \to 0} \frac{G(x, y+h) - G(x, y-h)}{F_Y(y+h) - F_Y(y-h)} \\ &= \frac{1}{f_Y(y)} \frac{dG(x,y)}{dy}, \; f_Y(y) > 0. \end{aligned}$$

The above leads to the following Bayes' theorem

$$f_{X|Y}(x|y) f_Y(y) = f_{Y|X}(y|x) f_X(x).$$

Example 0.2 *Let Y be a symmetric random variable with pdf f and $X = I_{\{Y>0\}}$. Let us compute conditional pmf/pdf.*

Note X is Bernoulli $(\frac{1}{2})$ and $f(y) = f(-y)$, since Y is symmetric.

$$\begin{aligned} G(0,y) &= P\big(X = 0, Y \leq y\big) \\ &= P\big(Y \leq 0, Y \leq y\big) \\ &= \begin{cases} \int_{-\infty}^{y} f(z)dz & \text{if} \quad y \leq 0 \\ \frac{1}{2} & \text{if} \quad y > 0. \end{cases} \end{aligned}$$

Similarly

$$\begin{aligned} G(1,y) &= P\big(X = 1, Y \leq y\big) \\ &= P\big(Y > 0, Y \leq y\big) \\ &= \begin{cases} 0 & \text{if} \quad y \leq 0 \\ \int_{0}^{y} f(z)dz & \text{if} \quad y > 0. \end{cases} \end{aligned}$$

Now

$$\begin{aligned}
\frac{dG(0,y)}{dy} &= \begin{cases} f(y) & \text{if} \quad y < 0 \\ 0 & \text{if} \quad y > 0, \end{cases} \\
\frac{dG(1,y)}{dy} &= \begin{cases} 0 & \text{if} \quad y < 0 \\ f(y) & \text{if} \quad y > 0. \end{cases}
\end{aligned}$$

Hence

$$\begin{aligned}
f_{Y|X}(y|x) &= 2\frac{dG(x,y)}{dy} \\
&= \begin{cases} 2f(y) & \text{if} \quad y < 0, x = 0 \text{ or } y > 0, x = 1 \\ 0 & \text{if} \quad \text{otherwise}, \end{cases}
\end{aligned}$$

Also for $f_Y(y) > 0$,

$$\begin{aligned}
f_{X|Y}(x|y) &= \frac{1}{f_Y(y)}\frac{dG(x,y)}{dy} \\
&= \frac{1}{f(y)}\frac{dG(x,y)}{dy} \\
&= \begin{cases} 1 & \text{if} \quad y < 0, x = 0 \text{ or } y > 0, x = 1 \\ 0 & \text{if} \quad \text{otherwise}, \end{cases}
\end{aligned}$$

Example 0.3 *Consider the random experiment given as follows.*

- *Pick a point at random from the interval* $(0,\ 1)$ *say* x
- *Now pick another point at random from the interval* $(0,\ x)$.

Find how is the second point distributed.

Let X *denote the position of the first point and* Y *denote the position of the second point.*

Then X *be uniform random variable over* $(0,\ 1)$ *and is given by*

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise}. \end{cases}$$

Note that the pdf of Y *given* $X = x$ *is* $U(0,\ x)$, *i.e.*

$$f_{Y\,|\,X}(y\,|\,x) = \begin{cases} \frac{1}{x} & 0 < y < x < 1 \\ 0 & \text{otherwise}. \end{cases}$$

Therefore,

$$f(x,\, y) \;=\; f_X(x) f_{Y\,|\,X}(y\,|\,x) \;=\; \begin{cases} \dfrac{1}{x} & 0 < y < x < 1 \\ 0 & \text{otherwise}\,. \end{cases}$$

Hence

$$f_Y(y) \;=\; \int_{-\infty}^{\infty} f(x,\, y)dx \;=\; \int_y^1 \frac{1}{x} dx \;=\; -ln\, y,\, 0 < y < 1\,.$$

0.4 Conditional probabilities

Using conditional pmf /pdf, we define conditional probabilities as follows. A remark about the notations. If Y is a discrete random variable, we denote D_Y be the set of discontinuities of its distribution function F_Y. The meaning of $f_{Y|X}$ is understood from the context, i.e. if Y is discrete, then it denote the conditional pdf of Y given X and when Y is continuous, then $f_{Y|X}$ denote the conditional pmf given X.

Definition 0.4 *If X, Y are random variable. Then for $-\infty < a < b < \infty$, we define*

$$\begin{aligned} P(a \le Y \le b\,|\,X \;=\; x) \;&=\; \sum_{y \in D_Y \cap [a,\; b]} f_{Y|X}(y|x) \text{ if } Y \text{ is discrete} \\ &=\; \int_a^b f_{Y|X}(y|x)dy \text{ if } Y \text{ is continuous.} \end{aligned}$$

In the above, when Y is continuous, it is assumed that $f_{Y|X}(y|x)$ exists.

WARNING : When X is a continuous random variable, then $P\{X = x\} = 0$, hence the LHS above doesn't have the meaning using the definition of conditional probability. So NEVER WRITE the LHS as

$$\frac{P\{a \le Y \le b, X = x\}}{P\{X = x\}}.$$

Example 0.4 *Let (X, Y) be uniformly distributed on $B(0, R)$, i.e. the open ball centered at $(0, 0)$ with radius R. Evaluate $P\Big(Y > 0\Big|X = x\Big)$.*

Note that

$$f(x,y) \;=\; \frac{1}{\pi R^2}, 0 \le x^2+y^2 < R, \;= 0 \text{ other wise.}$$

Also for $-R < x < R$*,*

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\ &= \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2} dy \\ &= \frac{2\sqrt{R^2-x^2}}{\pi R^2}. \end{aligned}$$

For $0 < x^2+y^2 < R$*,*

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f(x,y)}{f_X(x)} \\ &= \frac{1}{2\sqrt{R^2-x^2}}. \end{aligned}$$

Hence for $-R < x < R$*,*

$$\begin{aligned} P\Big(Y > 0\Big|X = x\Big) &= \int_0^{\sqrt{R^2-x^2}} \frac{1}{2\sqrt{R^2-x^2}} dy \\ &= \frac{1}{2}. \end{aligned}$$

Example 0.5 *Let* X, Y *be i.i.d. exp*(λ)*. Compute* $P\big(X > 1|X+Y = 2\big)$*.*

Let f *denote the joint pdf of* X, Y *and* g *denote the joint pdf of* $X, Z = X+Y$*. Then*

$$f(x,y) \;=\; \lambda^2 e^{-\lambda(x+y)}, \; x,y > 0, \;=\; 0 \text{ otherwise.}$$

Recall that

$$g(u,v) \;=\; \frac{1}{|\det A|} f((u,v)A^{-1}),$$

when $(U,V) = (X,Y)A$ *and* f, g *denote respectively the pdf of* (X,Y) *and* (U,V)*. One can see that*

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \; A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

Using this, it follows (exercise) that

$$g(x,z) \;=\; \lambda^2 e^{-\lambda z}, 0 < x < z, \;=\; 0 \text{ otherwise.}$$

Hence using the definition of $f_{X|Z}$, we get for $0 < x < z$,

$$\begin{aligned}
f_{X|Z}(x,z) &= \frac{g(x,z)}{f_Z(z)} \\
&= \frac{\lambda^2 e^{-\lambda z}}{\int_0^z g(x,z)dx} \\
&= \frac{\lambda^2 e^{-\lambda z}}{z\lambda^2 e^{-\lambda z}} \\
&= \frac{1}{z}.
\end{aligned}$$

Also $f_{X|Z}(x|z) = 0$ otherwise. Now

$$\begin{aligned}
P(X > 1|X+Y=2) &= \int_1^\infty f_{X|Z}(x|2)dx \\
&= \int_1^2 \frac{1}{2}dx \;=\; \frac{1}{2}.
\end{aligned}$$