E2:202 Random Processes **Nov. 6, 2020**

Tutorial 5: Jensen's Inequality, Conditional Distributions and Expectation

Lecturer: Parimal Parag *TA: Arvind* *Scribes: Krishna Chaythanya KV*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

5.1 Jensen's Inequality

Recall the definition of a convex function.

Definition (Convex Function)**.** A real-valued function $f : \mathbb{R} \to \mathbb{R}$ is convex if for all $x, y \in \mathbb{R}$ and $\theta \in [0, 1]$,

$$f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y).$$

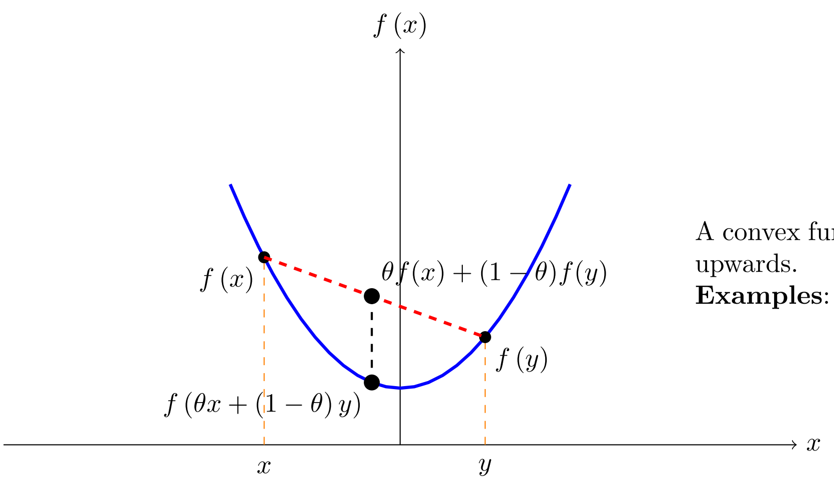


Figure 5.1: Convex Function

A convex function, hence looks like a cup opened upwards.
Examples: $f(x) = x^2$, e^x, $-\log x$, and so on.

A function $f : \mathbb{R} \to \mathbb{R}$ is *concave* if $(-f)$ is convex.

Remark. $f(x) = ax + b$, for $a, b \in \mathbb{R}$ is **both** concave and convex.

We now state the Jensen's inequality.

Theorem (Jensen's Inequality)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space. Further, let $f : \mathbb{R} \to \mathbb{R}$ be a convex function, and let X and $f(\cdot)$ be such that $\mathbb{E}[X] < \infty$ and $\mathbb{E}[f(X)] < \infty$. Then

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]).$$

We will now look at a straightforward proof of Jensen's inequality which relies on the following *equivalent* definition of convex functions.

Definition (Convex Function)**.** Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable at all $x \in \mathbb{R}$. Then f is convex iff for all $x, y \in \mathbb{R}$,

$$f(y) \geq f(x) + f'(x)(y - x).$$

Remark. Equivalently, if f is twice differentiable, its second derivative is non-negative for all $x \in \mathbb{R}$ iff f is convex. The statement in the definition above then follows from an application of Taylor's theorem.

We now prove Jensen's inequality.

Proof. Since we are given that f is convex (and assuming that f is differentiable), for all $\omega \in \Omega$ and for all $x \in \mathbb{R}$:

$$\begin{aligned} f(X(\omega)) &\geq f(x) + f'(x)(X(\omega) - x), \\ \Rightarrow \mathbb{E}[f(X)] &\geq f(x) + f'(x)\,\mathbb{E}[X - x]. \end{aligned}$$

Now, by choosing $x = \mathbb{E}[X]$,

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]).$$

□

Exercise 5.1. 1. Let P and Q be two probability distributions over a finite sample space Ω. Then, for a convex function f, such that $f(1) = 0$, the f-divergence of P from Q is defined as

$$D_f(P||Q) \triangleq \mathbb{E}_Q\left[f\left(\frac{P(X)}{Q(X)}\right)\right],$$

where $X : \Omega \to \mathfrak{X}$(finite). Assume $P(x), Q(x) > 0$, $\forall x \in \mathfrak{X}$. Show that

$$D_f(P||Q) \geq 0.$$

2. We have earlier seen the definition of the MGF of a RV, X

$$M_X(\lambda) \triangleq \mathbb{E}\left[e^{\lambda X}\right], \text{ for } \lambda \in \mathbb{R}.$$

Show that $M_X(\lambda) \geq \lambda\,\mathbb{E}[X]$, for $\lambda \in \mathbb{R}$.

5.2 Problems on Condition Distributions and Expectations

We will now look at a few problems on conditional distributions of discrete and continuous RVs.

Example 5.2. 1. Let Y be a Poisson RV with mean $\mu > 0$, and let Z be a geometrically distributed RV with parameter p such that $0 < p < 1$. Assume that Y and Z are independent.

(a) Find $\mathbb{P}(Y < Z)$

Solution. We will make use of the fact that $Y \perp Z$.

$$\begin{aligned}
\mathbb{P}(Y < Z) &= \sum_{y=0}^{\infty} \mathbb{P}(Z < y)\,\mathbb{P}(Y = y) && \text{[Law of total probability and } Y \perp Z\text{]} \\
&= \sum_{y=0}^{\infty} (1-p)^y \frac{e^{-\mu}\mu^y}{y!} && \text{[Show that } \mathbb{P}(Z > y) = (1-p)^y\ \forall y \geq 0.\text{]} \\
&= e^{-\mu} \sum_{y=0}^{\infty} \frac{(\mu(1-p))^y}{y!}, \\
&= e^{-\mu} e^{\mu(1-p)} = e^{-\mu p}. && \text{[Using Taylor's theorem.]}
\end{aligned}$$

□

(b) Find $\mathbb{P}(Y = i \mid Y < Z)$, for $i \geq 0$.

Solution. Fix $i \geq 0$. Then

$$\begin{aligned}
\mathbb{P}(Y = i \mid Y < Z) &= \frac{\mathbb{P}(Y = i, Y < Z)}{\mathbb{P}(Y < Z)} \\
&= \frac{\mathbb{P}(Y = i, Z > i)}{\mathbb{P}(Y < Z)} \\
&= \frac{\left(\frac{e^{-\mu}\mu^i}{i!}\right)(1-p)^i}{e^{-\mu p}} \\
&= \frac{e^{-\mu(1-p)}(\mu(1-p))^i}{i!}.
\end{aligned}$$

Thus, conditioned on $\{Y < Z\}$, Y is Poisson distributed with parameter $\mu(1-p)$. □

(c) Calculate $\mathbb{E}[Y \mid Y < Z]$

Solution. In the previous part, we showed that the conditional distribution of Y, given that $\{Y < Z\}$ is Poisson. Hence,

$$\mathbb{E}[Y \mid Y < Z] = \mu(1-p).$$

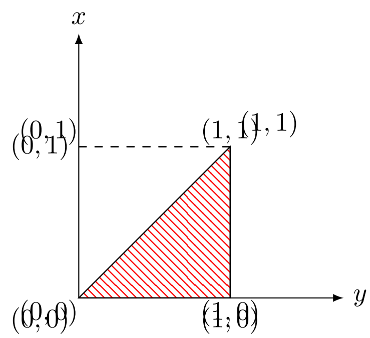
□

2. Suppose that RVs X and Y have the joint pdf

$$f_{X,Y}(x, y) = \begin{cases} 4x^2, & 0 < y < x < 1, \\ 0, & \text{o.w.} \end{cases}$$

(a) Find $\mathbb{E}[XY]$.

Solution. The region in $\mathbb{R}^2$ where the joint density is non-zero is depicted in the figure below.



From the density given, we get that

$$\begin{aligned}\mathbb{E}[XY] &= \int_0^1 \int_0^x (xy)\left(4x^2\right) dy dx \\ &= \int_0^1 4x^2 \left(\int_0^x y dy\right) dx \\ &= \int_0^1 4x^2 \cdot \frac{x^2}{2} = 2/5.\end{aligned}$$

□

(b) Compute $f_Y(y)$.

Solution. From the structure of the density function, we observe that $f_Y(y) > 0$ for $0 < y < 1$. Fix y s.t. $0 < y < 1$. Then, $x \in [y, 1]$, such that $f_{X,Y}(x, y) > 0$. Hence,

$$\begin{aligned}f_Y(y) &= \int_{x=y}^1 f_{X,Y}(x, y) dx \\ &= \int_{x=y}^1 4x^2 = \frac{4}{3}\left(1 - y^3\right), \text{ for } y \in (0, 1).\end{aligned}$$

Besides, $f_Y(y) = 0$, for $y \notin (0, 1)$. □

(c) Compute $f_{X|Y}(x \mid y)$.

Solution. Note that $f_{X|Y}(x \mid y)$ is defined only for $0 < y < 1$. Further,

$$\begin{aligned}f_{X|Y}(x \mid y) &= \frac{f_{X,Y}(x, y)}{f_Y(y)} \\ &= \begin{cases} \frac{3x^2}{1-y^3}, & x \in [y, 1], \\ 0, & \text{o.w.}\end{cases}\end{aligned}$$

□

(d) Compute $\mathbb{E}\left[X^2 \mid Y = y\right]$ for $0 < y < 1$ and thereby write down $\mathbb{E}\left[X^2 \mid Y\right]$.
(**Left as an exercise**)

Exercise 5.3. Let (X, Y) be uniformly distributed over the triangle with co-ordinates $(0,0), (1,0)$, and $(2,1)$.

1. What is the value of the joint pdf inside the triangle?
2. Find the marginal density of X, $f_X(x)$ for all $x \in \mathbb{R}$.
3. Find the conditional density function $f_{Y|X}(y \mid x)$ for all feasible values of x and y.
4. Calculate the conditional expectation $\mathbb{E}[Y \mid X = x]$.

We now proceed to a computational problem based on the law of iterated expectations.

Example 5.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let Y be a geometrically distributed RV with parameter $p \in (0,1)$.

$$\text{We note that } \mathbb{P}(Y < k) = (1-p)^k,\ k \in \{1,2,3,\ldots\}.$$

Let $X = 1_{\{A\}}$, where $A \triangleq \{\omega \in \Omega : Y(\omega) = 1\}$. Now,

$$\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}[\mathbb{E}[Y \mid X]] \\
&= \sum_{x=0}^{1} \mathbb{E}[Y \mid X = x]\,\mathbb{P}_X(x) \\
&= (1-p)\,\mathbb{E}[Y \mid X = 0] + p\,\mathbb{E}[Y \mid X = 1] \quad \textbf{Why?} \qquad (5.1)
\end{aligned}$$

$$\text{Now, } \mathbb{E}[Y \mid X = 1] = \mathbb{E}[Y \mid Y = 1] = 1.$$

$$\begin{aligned}
\text{Further, } \mathbb{E}[Y \mid X = 0] &= \mathbb{E}[Y \mid Y > 1] \\
&= 1 + \mathbb{E}[Y - 1 \mid Y < 1].
\end{aligned}$$

Now, we claim that $\mathbb{E}[Y - 1 \mid Y > 1] = \mathbb{E}[Y]$. This is because

$$\mathbb{P}(Y - 1 > k | Y > 1) = (1-p)^k. \quad \text{[Memoryless property]}$$

Hence, substituting in Eq. (5.1), we get

$$\begin{aligned}
\mathbb{E}[Y] &= p + (1-p)(1 + \mathbb{E}[Y]) \\
\Rightarrow \mathbb{E}[Y] &= \frac{1}{p}.
\end{aligned}$$

Example 5.5. Suppose that $X_1, X_2, \ldots$ are i.i.d. RVs with $\mathbb{E}[X_1] < \infty$. Suppose that N is another RV independent of X_n for all $n \in \mathbb{N}$ such that $N \in \{1, 2, \ldots\}$ and $\mathbb{E}[N] < \infty$. Then show that

$$\mathbb{E}\left[\sum_{n=1}^{N} X_n\right] = \mathbb{E}[N]\,\mathbb{E}[X_1].$$

(This is an important example, a modified version of which you will encounter again, when you study processes.)

Solution. Let $S_N \triangleq \sum_{n=1}^{N} X_n$. Note that the number of terms in the sum is a RV! We know that

$$\mathbb{E}[S_N] = \mathbb{E}[\mathbb{E}[S_N | N]]. \qquad (5.2)$$

Further,

$$\mathbb{E}\left[S_N|N=n\right] = \mathbb{E}\left[\sum_{i=1}^{N}|N=n\right],$$
$$= \mathbb{E}\left[\sum_{i=1}^{n} x_i\right],$$

where the last inequality is valid because $N \perp (X_i)_{i\in\mathbb{N}}$ and using the fact that $(X_i)_{i\in\mathbb{N}}$ are i.i.d., we have $\mathbb{E}\left[S_N|N=n\right] = n\mathbb{E}\left[X_1\right]$. Hence,

$$\mathbb{E}\left[S_N|N\right] = N\mathbb{E}\left[X_1\right].$$

Now, from Eq. (5.2), we have $\mathbb{E}\left[S_N\right] = \mathbb{E}\left[N\mathbb{E}\left[X_1\right]\right]$ and hence,

$$\mathbb{E}\left[S_N\right] = \mathbb{E}\left[N\right]\mathbb{E}\left[X_1\right]$$

□