Probability 1 - Lecture Notes

Philip Warton

November 13, 2020

1 Markov Inequality

Suppose there is a distribution for which we don't know the probability mass function, and we do not know the variance, but we do know it's expectation, $E[x]$. What can we say about that probability? Can we bound it?

Theorem 1.1 (Markov Inequality)**.** *If X is a random variable that takes only non-negative values, then for any $\alpha > 0$,*

$$P(X \leq \alpha) \leqslant \frac{E[x]}{\alpha}$$

Proof.

$$P(X \geq \alpha) = \sum_{k:k\geq\alpha} p(\alpha) \leq \sum_{k:k\geq\alpha} \frac{k}{\alpha} p(k) = \frac{1}{\alpha} \sum_{k:k\geq\alpha} k \cdot p(k) \leq \frac{1}{\alpha} \sum_{k:k\geq 0} k \cdot p(k) = \frac{E[X]}{\alpha}$$

□

Note that this would likely work under integration for a continuous random variable.

Theorem 1.2 (Chebyshev Inequality)**.** *If X is a random variable with a finite mean μ and variance, then for any $\kappa > 0$,*

$$P(|X - \mu| \geq \kappa\sigma) \leq \frac{1}{\kappa^2}$$

2 Continuous Random Variables

Definition 2.1. We say that X is a continuous random variable if there exists a nonnegative function $f(x)$ defined for all real x such that for any $a \leq b$

$$P(a \leq X \leq b) = \in_a^b f(x)dx$$

Such a function $f(x)$ is the probability density function of X $\boxed{\text{Figure 1}}$.

First notice that the prboability density function must be non-negative, because it is impossible to have a negative probability by definition axiomatically. There are some properties of these functions that we will enumerate now:

(i) $\displaystyle\int_{-\infty}^{\infty} f(x)dx = P(-\infty < X < \infty) = 1$

(ii) $P(X = a) = \displaystyle\int_a^a f(x)dx = 0 \forall a \in \mathbb{R}$

(iii) $P(a < X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b) = \displaystyle\int_a^b f(x)dx$

We can restate this definition by saying, $f(x)$ is a probability density function $\Leftrightarrow$ $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)dx = 1$. Even though $P(X = a) = 0$ for every real number a, since the real numbers are uncountable, we do not violate any of our axioms of probability. Since $P(S) = 1$ for any sample space S, it follows that $P(-\inf \leq X \leq \inf) = 1$.

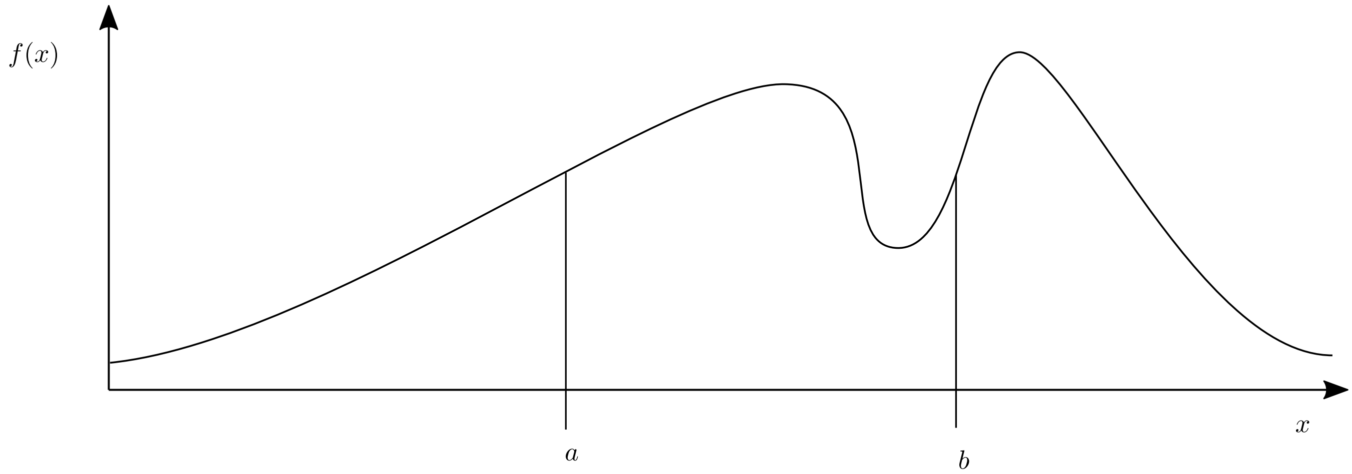

Figure 1: Probability Density Function

Definition 2.2. Let X be a continuous random variable with density function $f(x)$. Then its expectation is

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

We carry some similar properties over from discrete expectation. Firstly,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

The $\boxed{\text{Markov Inequality}}$ also will hold for continuous random variables. That is,

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

Proof. Let $\alpha > 0$. For every $\alpha \leq x < \infty, 1 \leq \frac{x}{\alpha}$. Then we say

$$P(X \geq \alpha) = \int_{\alpha}^{\infty} f(x)dx \leq \int_{\alpha}^{\infty} \frac{x}{\alpha} f(x)dx$$

The right hand side is bounded by $\frac{1}{\alpha}\int_0^{\infty} xf(x)dx = \frac{1}{\alpha}E[X]$. □

Finally the $\boxed{\text{Chebyshev Inequality}}$ also will hold:

$$P(|X - \mu| \geq \kappa) \leq \frac{Var(x)}{\kappa}$$

The proof of the Chebyshev Inequality does not change from the proof in the discrete case.

2.1 Exponential Random Variable

Let us take the example of the following function:

$$f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

We know that this function will integrate to 1 over $\mathbb{R}$. Scaling, this function by λ we get another probability density function.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

We obtain the same exact area, so we still have a valid probability density function so long as $\lambda > 0$. This is called an exponential random variable. It is a continuous analogue to the geometric random variable in the discrete case. Then it also carries the property of memorylessness, which means that $P(X > a + b | X > a) = P(X > b)$, for any $a, b \geq 0$. Generally this is because after shifting our start point to a, and normalizing the distribution, we simply get the same function again. However, we must prove this more rigorously.

Proof. For any $a > 0$, we can first compute the probability that $X > a$.

$$P(X > a) = \int_a^{\inf} \lambda e^{-\lambda x} dx = (-e^{-\lambda x})_a^{\inf} = e^{-\lambda a}$$

Then the conditional probability can be computed as follows:

$$P(X > a + b | X > a) = \frac{P(X > a + b)}{P(X > a)} = \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} = e^{-\lambda b}$$

□

We proved in class that memorylessness is unique to the exponential random variable. To find the expectation of such an exponential variable let $\lambda > 0$. We say that

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_0^{\infty} \lambda x \cdot e^{-\lambda x} dx$$

Then we must use integration by parts, which gives us $-xe^{-\lambda x}\Big|_0^{\infty} - \int_0^{\infty} \lambda e^{-\lambda x} = \frac{1}{\lambda}$.

2.2 Uniform Random Variable

Consider a real interval $[\alpha, \beta] : \alpha < \beta$. Let X be a continuous random variable with density funciton

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

With this funciton X is a uniform random variable over the interval $[\alpha, beta]$.

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{a} 0 \cdot dx + \int_{\alpha}^{\beta} \frac{dx}{\beta - \alpha} + \int_{\beta}^{\infty} 0 \cdot dx = \left[\frac{x}{\beta - \alpha}\right]_{\alpha}^{\beta} = 1$$

2.3 Normal (Gaussian) Random Variable

X is a normal random variable with parameters μ and σ^2 if its density function its

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

Let $\mathcal{N}(\mu, \sigma^2)$ denote a normal distribution with parameters μ and σ^2. The expectation of the Normal Random Variable is σ, this can be shown by computing the integral $\int_{-\infty}^{\infty} x \cdot f(x) dx$.

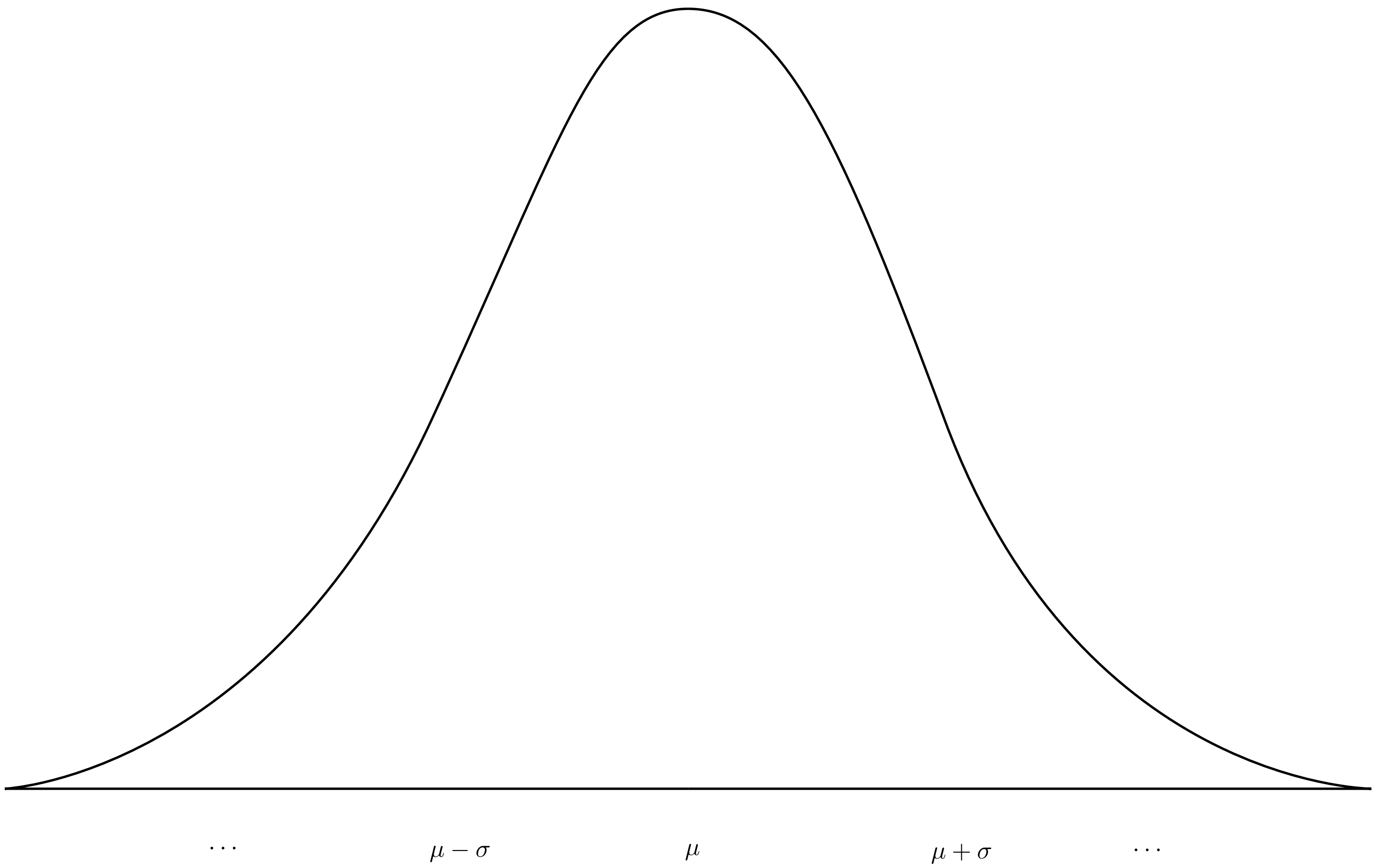

Figure 2: Probability Density Function of Normal Distribution