EEECS 126

Albert Ye

September 23, 2023

1 Probability Space

1.1 Definition

Essentially from 70. Events happen with some probability in a larger probability space that contains all events that can happen.

1.2 Axioms of Probability

Proposition 1 (Axioms) 1. (Positivity) $P(\omega > 0)$ for any event ω in probability space Ω.

2. (Totality) In any sample space Ω, $P(\Omega) = 1$.
3. (Additivity) If $A_1, A_2, \ldots, A_n$ are independent, then

$$\sum_{i=1}^{n} A_i = \bigcup_{i=1}^{n} A_i.$$

From just this, we can get some useful information, such as the union bound.

Theorem 2 (Union Bound)

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

The proof is left as an exercise to the student, probably in the homework.

1.3 σ-algebra

Definition 3 (σ-algebra)

Given a sample space Ω, a set $\mathcal{F} \subseteq 2^{\Omega}$ is a σ-algebra if:

1. $\Omega \in \mathcal{F}$
2. If any event A is in $\mathcal{F}$, then its complement $\Omega \setminus A$ is also in $\mathcal{F}$.
3. For countably many events $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}$, their union $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

2 Conditional Probability

2.1 Definition

2.2 Total Probability

2.3 Bayes' Rule

2.4 Continuous Bayes

3 It Depends

3.1 Independence / (Un)correlation

3.2 Conditional Expectation

Notice that $E[X|Y]$ is a random variable, but $E[X|Y = y]$ is a number. We can call $E[X|Y]$ a function $g(Y)$, where then $E[X|Y = y] = g(y)$ is just a value in the function.

3.3 Iterated Expectation

4 Distributions

4.1 Joint Distribution

4.2 Marginal Distribution

4.3 Derived Distribution

5 Random Variables

5.1 Discrete

5.1.1 Bernoulli

- PMF: $p_X(k) = \begin{cases} p & k = 1 \\ 1 = p & k = 0 \end{cases}$
- Expected value: p
- Variance: $p(1-p)$.

5.1.2 Binomial

- PMF: $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ over all $k \in 0, 1, \ldots, n$.
- Expected value: np
- Variance: $np(1-p)$.

Run a Bernoulli test n times, find how many are positive.

5.1.3 Geometric

- PMF: $p_X(k) = (1-p)^{k-1}p$, for $k = 1, 2, \ldots$.
- Expected value: $\frac{1}{p}$
- Variance: $\frac{1-p}{p^2}$.

Here, each trial has a p probability of success, and we want to find the # of trials until one success.

5.1.4 Poisson

- PMF: $p_X(k) = \frac{\lambda^k (e^{-\lambda})}{k!}$.
- Expected value: λ
- Variance: λ

Used to simulate arrivals, I guess. More useful later, with Poisson processes.

5.2 Continuous

5.2.1 Uniform

5.2.2 Exponential

5.2.3 Gaussian

5.2.4 Joint Gaussian

The main tips for Joint Gaussian are to approach it as a sort of vectorized Gaussians over a certain number N of dimensions. Most of the addition / whatever operations in a Gaussian can be remodeled as a Joint Gaussian.

6 Moment Generating Functions

Definition 4

The **moment generating function** (also known as a transform) associated with a RV X, is a function $M_X(s)$ of a scalar parameter s defined by $M_X(s) = E(e^{sX})$.

the simpler notation $M(S)$ can be used whenever the underlying random variable X is clear from context. In more detail, when X is a discrete random variable, the corresponding MGF is given by

$$M(s) = \sum_x e^{sx} p_X(x).$$

Analogously, when continuous, we just replace the summation with an integral to get

$$M(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x) dx.$$

Just an example so that I know what the reference is here:

Example 5 (Discrete Example)

Let

$$p_X(x) = \begin{cases} \frac{1}{2} & x = 2 \\ \frac{1}{6} & x = 3 \\ \frac{1}{3} & x = 5. \end{cases}$$

Then the corresponding transform is

$$M(s) = E(e^{sx}) = \frac{1}{2} + \frac{1}{6}e^{3s} + \frac{1}{3}e^{5s}.$$

Example 6 (Continuous Example)

Let X be an exponential RV with parameter λ:

$$f_X(x) = \lambda e^{-\lambda x} \qquad x \geq 0.$$

Then,

$$\begin{aligned} M(s) &= \lambda \int_0^{\infty} e^{sx} e^{-\lambda x} dx \\ &= \lambda \int_0^{\infty} e^{(s-\lambda)x} dx \\ &= \lambda \left(\frac{e^{(s-\lambda)x}}{s-\lambda} \right|_0^{\infty} \\ &= \frac{\lambda}{\lambda - s}. \end{aligned}$$

Notice, in above examples, that MGF is a **function** of parameter s, and not a number. We can also find MGF's for functions of X:

Proposition 7 (MGF of Linear Function of RV)

Let $Y = aX + b$. Then,

$$M_Y(s) = E(e^{s(aX+b)}) = e^{sb} E(e^{saX}) = e^{sb} M_X(sa).$$

From our previous example, we see that $M_X(s) = \frac{1}{1-s}$ where X is the exponential distribution

6.1 Moments

Now that we've established what a moment generating function is, now it's time to understand what is being generated.

Let's do a generic MGF

$$M(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x)dx.$$

Now, we take the derivative of this.

$$\begin{aligned}\frac{d}{ds}M(s) &= \frac{d}{ds}\int_{-\infty}^{\infty} e^{sx} f_X(x)dx \\ &= \int_{-\infty}^{\infty} \frac{d}{ds}e^{sx} f_X(x)dx \\ &= \int_{-\infty}^{\infty} xe^{sx} f_X(x)dx.\end{aligned}$$

When $s = 0$, we have that this evaluates to $\int_{-\infty}^{\infty} x f_X(x)dx = E(X)$. If we differentiate n times, then we will get

$$\left(\frac{d^n}{ds^n}M(s)\right|_{s=0} = \int_{-\infty}^{\infty} x^n f_X(x)dx = E(X^n).$$

6.2 Inversion

Proposition 8 (Inversion Property)

The MGF $M_X(s)$ associated with an RV X uniquely determines the CDF of X, assuming that $M_X(s)$ is finite for all s in some interval $[-a, a]$ for positive a.

6.3 Sum of Independent Random Variables

Proposition 9

Addition of independent random variables corresponds to multiplication of transforms.

Proof. Let $Z = X + Y$. $M_Z(s) = E(e^{sZ}) = E(e^{s(X+Y)}) = E(e^{sX}e^{sY})$. Since X, Y are independent, e^{sX} and e^{sY} are independent random variables for any fixed s. Thus, $E(e^{sX}e^{sY}) = E(e^{sX})E(e^{sY}) = M_X(s)M_Y(s)$. □

We can further extend this; if $X_1, \ldots, X_n$ is a collection of independent random variables and $Z = X_1 + \cdots + x_n$, then $M_Z(s) = M_{X_1}(s)\cdots M_{X_n}(s)$.

7 Concentration Inequalities

Theorem 10 (Markov's Inequality)

Theorem 11 (Chebyshev's Inequality)

8 Modes of Convergence

8.1 Pointwise

Definition 12 (Pointwise Convergence)
Fix $\omega \in \Omega$, $\{X_n(\omega)\}_{n=1}^{\infty}$ converges **pointwise** if it becomes a real-valued sequence.

Usually, people don't use this because of reasons highlighted in 104.

8.2 Almost Sure

Definition 13 (Almost Sure Convergence)
$\{x_n\}_{n=1}^{\infty}$ converges **almost surely** to X if $P(\{\omega : \omega \in \Omega, \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$.

This gets rid of ω with probability 0. If you find an ω such that convergence doesn't hold, it's fine as long sa $P(\omega) = 0$.

8.2.1 Checking for Almost Sure Convergence

There are a couple ways to check if some sequence converges almost surely.

8.3 In Probability

This is a weaker bound for convergence than almost sure convergence.