1 - Distribution 1

> Statistics
>
> Statistics is a mathematical way of collection, analysis, interpretation, and visualization of data

Review of Probability

This course will focus on the statistical side, whilst MATH 411 focused on the probability theory. We will now recall some basics on probability:

> Probability Space
>
> A Probability Space is is a 3-tuple $(\Omega, \mathcal{F}, P)$, where Ω is a set, $\mathcal{F} \subset P(\Omega)$ is a Sigma-Algebra, and $P : \Omega \mapsto [0, \infty]$ is a measure such that $P(\Omega) = 1$.

> Event (Probability)
>
> Given a Probability Space $(\Omega, \mathcal{F}, P)$, an Event is any $E \in \mathcal{F}$.

Recall the concepts of Conditional Probability and Independence.

Random Variables

Once again reviewing definitions

> Random Variable
>
> Given a probability space, a Random Variable is a measurable function $X : \Omega \mapsto E$, where E is a measurable space. In the standard case we have $E = \mathbb{R}$.
>
> The probability that X takes on a value in a measureable set $S \subset E$ is written as
>
> $$P(X \in S) = P(\{\omega \in \Omega : X(\omega) \in S\})$$

We will be, at least for now, studying the $E = \mathbb{R}$ case.

The two main types of random variables we will be studying are discrete random vairbales and continuous random variables.

> Cumulative Distribution Function
>
> The Cumulative distribution function or CDF of a $\mathbb{R}$-valued random variable is the function $F : \mathbb{R} \mapsto [0, 1]$ defined as
>
> $$F(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\})$$
>
> the CDF characterizes the distriubtion of X (which is any probability of the form $P(X \in A)$ for $A \subset \mathbb{R}$).

However, we often employ the PMF/PDF to charcaterize the distribution.

> Probability Mass Function
>
> For discrete random variables, the PMF of X is a function $f : \mathbb{N} \mapsto [0, 1]$ such that
>
> $$f(x) = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

> Probability Density Function
>
> The PDF of a contiuous random variable X is a function $f : \mathbb{R} \mapsto [0,1]$$ that satisfies
>
> $$F(x) = \int_{-\infty}^{x} f(u)du$$
>
> Moreover $f(x) = F'(x)$

Review of Common Discrete Distributions

- Discrete Uniform Distribution
- Hypergeometric Distribution
- Binomial Distribution
- Poisson Distribution
- Geometric Distribution
- Negative Binomial Distribution
- Bernoulli Distribution

Review of Common Continuous Distributions

- Continuous Uniform Distribution
- Exponential Distribution
- Gamma Family
- Normal Distribution
- Beta Distribution

Random Vectors and Independence

> Random Vector
>
> A random vector $(X_1, \ldots, X_n)$ is a vector of random variables for which we have a joint probaability distribution. The joint CDF is defined as
>
> $$F(x_1, \ldots, x_n) = P(X_1 \le x_1, \ldots, X_n \le x_n)$$

The marginal distribution of a random variable can be obtained from the joint distribution. For example the marginal CDF of X_1 is given by

$$F_{X_1}(x_1) = \lim_{x_2 \to \infty} \cdots \lim_{x_n \to \infty} F(x_1, \ldots, x_n)$$

Now recall the definition of

> Independence (Probability)
>
> We say that random variables $(X_1 \ldots X_n)$ are independent if the joint distribution factors as the product of the marginals for all possible combinations of real-values arguments:
>
> $$F(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n$$

Independence in Statistics

A random sample is typically modeled as a collection of independent and identically distributed (i.i.d.) random variables
We write $X_1, \ldots, X_n \sim F$ to denote a random sample of size n from a distribution with cdf F. The joint CDF can be expressed as

$$F(x_1, \ldots, x_n) = F(x_1) \cdots F(x_n) = \prod_{i=1}^{n} F(x_i)$$

Similarly the joint PMF/PDF can be written as

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$$

Example: A Random Sample From a Bernoulli Distribution

Consider a random sample of size n from the Bernoulli distribution with parameter $p \in (0,1)$. Show that the joint PMF can be written only as a function that depends on the sample mean

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Solution

The joint PMF is:

$$P(X_1 = x_1, \ldots, X_n = x_n)$$

Since $(X_i)_{i=1}^{n}$ are i.i.d then this is equal to

$$\begin{aligned}
&= \prod_{i=1}^{n} f(x_i) \\
&= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\
&= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \\
&= (p(1-p)^{n-1})^{\sum_{i=1}^{n} x_i} \\
&= (p(1-p)^{n-1})^{n\bar{X}} \\
&= (\frac{p}{1-p})^{\bar{X}}
\end{aligned}$$

Question

Should now note I am confused -- as always -- as to the roles of X_i and x_i. The former is a measurable function $\Omega \mapsto \mathbb{R}$. The latter is a simply a value in $\mathbb{R}$. Why was I able to change $\sum_{i=1}^{n} x_i$ to $\sum_{i=1}^{n} X_i$? Because they are equal retard. By the very premise of the question or right above where you wrote $P(X_1 = x_1, \ldots, X_n = x_n)$.

Expectation of Random variables

> Expectated Value
>
> The expectation of a random variable in a given probability space $(\Omega, \mathcal{F}, P)$ is essentially a weighted average. More rigorously:
>
> $$E(X) = \int_{\Omega} X(\omega)\, dP(\omega) = \int_{\Omega} X\, dP$$
>
> Note the trivial application to the standard case in $\mathbb{R}$ as well as the discrete case.
>
> **Note**
>
> μ is often used to denote $E(X)$

We can also calculate the expectations of functions of random variables and vectors by integrating using the appropriate PDF/PMF: given $X : \Omega \mapsto E$ and $g : E \mapsto M$, where M is a measurable space

$$E(g(X)) = \int_{\Omega} g(X(\omega))\, dP(\omega) = \int_{\Omega} g(X)\, dP$$

We will now recall some important quantities:

Random Variables

The r-th moment of a random variable X is

$$\mu_r = E(X^r) = \int_\Omega X^r dP$$

Note in the definition below that Variance is simply the 2nd-moment of $X - E(X)$:

Variance

The variance of a random variable X is given by

$$\sigma^2 = Var(X) = E((X - E(X))^2)$$

Covariance

The covariance of two random variables, X and Y, is given by

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$

Recall the following properties of expectations:

- Linearity: $E(aX + Y) = aE(X) + E(Y)$ for any $a \in \mathbb{R}$
- Variance Identity: $Var(X) = E(X^2) - E(X)^2$
- Covariance Identity: $Cov(X, Y) = E(XY) - E(X)E(Y)$
- $Var(aX + b) = a^2 Var(X)$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
- $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
- If X and Y are Independent, then $E(XY) = E(X)E(Y) \implies Cov(X, Y) = 0$
- And many more (Check Chapter 3 of the textbook)

Sample Mean and Sample Variance

Sample Mean

Consider a random sample of i.i.d. random variables $X_1 \dots X_n \sim F$. The sample mean is defined as the random variable $\bar{X}$ such that

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Sample Variance

Consider a random sample of i.i.d. random variables $X_1 \dots X_n \sim F$. The sample variance is defined as the random variable S^2 such that

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Where $\bar{X}$ is the sample mean.

Note we will sometimes write the subscript $\bar{X}_n$ or S_n^2 to emphasize the dependence on the sample size n.

⚠ Proposition

Given a random sample of i.i.d. random variables $X_1 \dots X_n \sim F$ the expectation and variance of the sample mean are given by

$$E(\bar{X}) = \mu = E(X_i)$$

and

$$Var(\bar{X}) = \frac{\sigma}{n^2}$$

Fruthermore, the expectation and variance of the sample variance are given by

$$E(S^2) = \sigma^2$$

and

$$Var(S^2) = \frac{1}{n}(\mu_4 - \frac{n-3}{n-4}\sigma^4)$$

Proof

Will only prove the first claim.

$$E(\bar{X}) = \frac{1}{n}\sum_i E(X_i) = \frac{1}{n}n\mu = \mu$$

We note the claim on $Var(\bar{X})$ was done in homework #1. Simply recall that $X_1 \dots X_n \sim F$ are independent and identically distributed random variables $\implies Cov(X_i, X_j) = 0$

Moment Generating function

Recall the r-th moment of a random variable.

> Moment Generating Function
>
> Given a random variable X, the function
>
> $$M_x(t) = E(e^{tX})$$
>
> is called the moment generating function, or mgf, for X, assuming it is finite for all t in some open interval containing zero

Let X be a random variable with mgf $M_x(t)$. The moments of X can be obtained by

$$E(X^r) = \frac{d^r M_x(t)}{dt^r}\Big|_{t=0}$$

Properties of Moment Generating Functions

Linear Combinations of Random Variables via MGF

Asymptotic Theory

In statistics we are often interested in functions $g_n : \mathbb{R}^n \mapsto \mathbb{R}$ that maps a random variable of size n into a real-valued random variable T_n:

$$T_n = g_n(X_1, \dots, X_n)$$

For example take the sample mean and sample variance
We often need to know the distribution of T_n to make probabilistic statements about it. In some cases deriving the exact distribution of T_n is possible (for example the sample mean of a normal distribution).
When the exact distribution is not available, we often need to rely on large sample theory.
We are interested in asymptotic properties of T_n, specifically:

- Does the sequence $(T_n)_{n\in\mathbb{N}}$ converge to a meaningful limit?
- What is the distribution of T_n as $n \to \infty$?

> Convergence in Measure
>
> Let (X, Σ, μ) be a be a measure space. Let $f, (f_n)_{n \in \mathbb{N}} : X \mapsto \mathbb{R}$ be measurable functions. The sequence f_n is said to converge in measure, $f_n \xrightarrow{\mu} f$ if $\forall \epsilon > 0$
>
> $$\lim_{n \to \infty} \mu(\{x \in X : |f(x) - f_n(x)| \geq \epsilon\}) = 0$$

In our case we use probability space $(\Omega, \mathcal{F}, P)$. If we now let $f = \alpha \in \mathbb{R}$, in other words let it be a constant, then we have convergence in probability

> Weak Law of Large Numbers
>
> **Theorem**
>
> Consider a random sample $(X_i)_{n \in \mathbb{N}} \overset{\text{i.i.d.}}{\sim} F$ with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ (both finite). Then the sample mean $\bar{X}_n$ converges in probability to μ. In other words
>
> $$\bar{X}_n \xrightarrow{P} \mu$$
>
> ***Proof***
>
> TODO

Example: WLLN and Poisson Distribution

> Convergence in Distribution
>
> We say that a sequence of random variables $T_n \sim F_n$ converges in distribution to a random variable $T \sim F$. denoted by $T \xrightarrow{\mathcal{D}} T$ if $\forall x \in \mathbb{R}$:
>
> $$\lim_{n \to \infty} F_n(x) = F(x)$$

> Central Limit Theorem
>
> **Theorem**
>
> Let $(X_i)_{i \in \mathbb{N}} \overset{\text{i.i.d.}}{\sim} F$ be a random sample, with $E(X) = \mu$ and $Var(X) = \sigma^2$ (both finite). Then
>
> $$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{\mathcal{D}} Z \sim N(0, 1)$$
>
> In other words it converges in distribution to a normal random variable with parameters μ and σ^2
>
> ***Proof***
>
> TODO

Example: CLT and Poisson Distribution

Slide 33

Exponential Family

> Exponential Family
>
> A distribution is said to belong k-parameter exponential family if it has a PMF/PDF of the form:
>
> $$f(x;\theta_1,\dots,\theta_n) = c(\theta_1,\dots,\theta_n)h(x)\exp\left(\sum_{j=1}^{k} w_j(\theta_j)t_j(x)\right)$$
>
> Note when $k = 1$ we have the one parameter exponential family of the form:
>
> $$f(x,\theta) = c(\theta)h(x)\exp\Big(w(\theta)t(x)\Big)$$

Properties of the Normal Distribution