Week 7

Statistics 251

Zhongjian Wang

Department of Statistics
The University of Chicago

Where are we?

Sums of independent r.v.

Expectation of Sums

Covariance

Correlation

Sum of two independent r.v.

Given X, Y independent, continous r.v. with density function f_X, f_Y, the density function of $X + Y$ can be written as,

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)dy = \int_{-\infty}^{\infty} f_X(x)f_Y(a-x)dx$$

Sums of uniform distributions

If X_1, X_2 are independent identical uniform distributed on $(0, 1)$, what is the distribution of $X_1 + X_2$?

Continued..

If X_1, X_2, $\cdots$, are independent identical uniform distributed on $(0, 1)$. What is the expectation of N where

$$N = \min\{n : X_1 + X_2 + \cdots + X_n > 1\}$$

Let F_n denote cummulative distribution function of $X_1 + \cdots + X_n$. By Mathematical Induction, we first try to prove
$F_n(x) = x^n/n!, \quad 0 \leq x \leq 1.$

So $\mathbb{P}\{N > n\} = F_n(1)$

Sums of Normal distribution

Recall density function of a normal distribution with parameters (μ, σ^2) is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

Proposition: If X_1, X_2, $\cdots$, X_n are independent random random variables with respective parameters (μ_1, σ_1^2), (μ_2, σ_2^2), $\cdots$, (μ_1, σ_1^2),
then $X_1 + X_2 + \cdots + X_n$ is a normal random variable with parameters $(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)$.

Sums of Gamma distribution

Recall density function of a gamma distribution with parameters (α, λ) is given by

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Proposition: If X and Y are independent gamma random variables with respective parameters (s, λ) and (t, λ), then $X + Y$ is a gamma random variable with parameters $(s + t, \lambda)$.

• So α is refereed as shape parameter, λ is refereed as scale parameter.

You are also advised to read the relevant part (Poisson, binomial, geometric) on Ross book.

Where are we?

Sums of independent r.v.

Expectation of Sums

Covariance

Correlation

Revision: Expected Value of a function of a r.v.

If X and Y have a joint probability mass function $p(x, y)$, then

$$E[g(X, Y)] = \sum_y \sum_x g(x, y)p(x, y)$$

If X and Y have a joint probability density function $f(x, y)$, then

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)dxdy$$

Derivation

Suppose that $E[X]$ and $E[Y]$ are both finite and let $g(X, Y) = X + Y$. Then, in the continuous case,

$$
\begin{aligned}
E[X+Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+y) f(x,y) dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y) dy dx + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x,y) dx dy \\
&= \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy \\
&= E[X] + E[Y]
\end{aligned}
$$

What about $E\left[X_1 + \cdots + X_n\right]$?

Sample Mean

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having distribution function F and expected value μ.
Then $X_1, \ldots, X_n$ is said to constitute *a sample from the distribution F*.
The quantity

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

is called the sample mean.
Now what is $E[\bar{X}]$?

Note: when the distribution mean μ is unknown, the sample mean is often used in statistics to estimate it.

Example: A Summation Formula

Consider any nonnegative, integer-valued random variable X. If, for each $i \geq 1$, we define

$$X_i = \begin{cases} 1 & \text{if } X \geq i \\ 0 & \text{if } X < i \end{cases}$$

then $E[X] = \sum_{i=1}^{\infty} E\left(X_i\right)$

Note $E\left(X_i\right) = P\{X \geq i\}$, so

$$E[X] = \sum_{i=1}^{\infty} P\{X \geq i\}$$

a useful identity.

Example: Sorting elements

Suppose that n elements, $1, 2, \cdots, n$ must be stored in a computer in the form of an ordered list. Each unit of time, a request will be made for one of these elements i being requested, independently of the past, with known probability $P(i), i \geq 1$, $\sum_i P(i) = 1$.
What ordering minimizes the average position in the line of the element requested?

Suppose that the elements are numbered so that $P(1) \geq P(2) \geq \cdots \geq P(n)$. Let X denote the position of the requested element. Now, under any ordering, $O = i_1, i_2, \ldots, i_n$,

$$P_O\{X \geq k\} = \sum_{j=k}^{n} P\left(i_j\right)$$

Where are we?

Sums of independent r.v.

Expectation of Sums

Covariance

Correlation

Expected value of multiplication of functions of independent r.v.

If X and Y are independent, then, for any functions h and g,

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Continuous case proof:

Definition of Covariance

Recall definition of Var X:

The covariance between X and Y, denoted by Cov (X, Y), is defined by

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Alternative form

$$\mathsf{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

$$\Updownarrow$$

$$\mathsf{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Properties

(i) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$

(ii) $\text{Cov}(X, X) = \text{Var}(\mathrm{X})$

(iii) $\text{Cov}(aX, Y) = a\,\text{Cov}(X, Y)$

(iv) $\text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{m} Y_j\right) = \sum_{i=1}^{n} \sum_{i=1}^{m} \text{Cov}\left(X_i, Y_j\right)$

Variance of sum

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) + 2\sum_{i<j} \mathrm{Cov}\left(X_i, X_j\right)$$

Example: Sample Variance

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having expected value μ and variance σ^2, Llet $\bar{X} = \sum_{i=1}^{n} X_i/n$ be the *sample mean*. Then what is $\mathrm{Var}(\bar{X})$? The quantities $X_i - \bar{X}, i = 1, \ldots, n$, are called *deviations*, as they equal the differences between the individual data and the sample mean. The random variable

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

is called the *sample variance*. Find $\mathbb{E}\left[S^2\right]$.

Where are we?

Sums of independent r.v.

Expectation of Sums

Covariance

Correlation

Definition

The correlation of two random variables X and Y, denoted by $\rho(X, Y)$, is defined, as long as $\operatorname{Var}(X)\operatorname{Var}(Y)$ is positive, by

$$\rho(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$$

Then $-1 \leq \rho(X, Y) \leq 1$.

Example: Derivation and sample mean are uncorrelated

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having variance σ^2. Then

$$\mathrm{Cov}\left(X_i - \bar{X}, \bar{X}\right) = 0$$

Example: if $Y = a + bX$

Given mean and variance of X to be μ and σ^2, calculate $\rho(X, Y)$, where $Y = a + bX$.