18.440: Lecture 22

Joint distributions functions

Scott Sheffield

MIT

Outline

Distributions of functions of random variables

Joint distributions

Independent random variables

Examples

Outline

Distribution of function of random variable

- Suppose $P\{X \leq a\} = F_X(a)$ is known for all a. Write $Y = X^3$. What is $P\{Y \leq 27\}$?
- Answer: note that $Y \leq 27$ if and only if $X \leq 3$. Hence $P\{Y \leq 27\} = P\{X \leq 3\} = F_X(3)$.
- Generally $F_Y(a) = P\{Y \leq a\} = P\{X \leq a^{1/3}\} = F_X(a^{1/3})$
- This is a general principle. If X is a continuous random variable and g is a strictly increasing function of x and $Y = g(X)$, then $F_Y(a) = F_X(g^{-1}(a))$.
- How can we use this to compute the probability density function f_Y from f_X?
- If $Z = X^2$, then what is $P\{Z \leq 16\}$?

Outline

Outline

Distributions of functions of random variables

Joint distributions

Independent random variables

Examples

Joint probability mass functions: discrete random variables

- If X and Y assume values in $\{1, 2, \ldots, n\}$ then we can view $A_{i,j} = P\{X = i, Y = j\}$ as the entries of an $n \times n$ matrix.
- Let's say I don't care about Y. I just want to know $P\{X = i\}$. How do I figure that out from the matrix?
- Answer: $P\{X = i\} = \sum_{j=1}^{n} A_{i,j}$.
- Similarly, $P\{Y = j\} = \sum_{i=1}^{n} A_{i,j}$.
- In other words, the probability mass functions for X and Y are the row and columns sums of $A_{i,j}$.
- Given the joint distribution of X and Y, we sometimes call distribution of X (ignoring Y) and distribution of Y (ignoring X) the **marginal** distributions.
- In general, when X and Y are jointly defined discrete random variables, we write $p(x, y) = p_{X,Y}(x, y) = P\{X = x, Y = y\}$.

Joint distribution functions: continuous random variables

- Given random variables X and Y, define $F(a, b) = P\{X \leq a, Y \leq b\}$.
- The region $\{(x, y) : x \leq a, y \leq b\}$ is the lower left "quadrant" centered at (a, b).
- Refer to $F_X(a) = P\{X \leq a\}$ and $F_Y(b) = P\{Y \leq b\}$ as **marginal** cumulative distribution functions.
- Question: if I tell you the two parameter function F, can you use it to determine the marginals F_X and F_Y?
- Answer: Yes. $F_X(a) = \lim_{b\to\infty} F(a, b)$ and $F_Y(b) = \lim_{a\to\infty} F(a, b)$.

Joint density functions: continuous random variables

- Suppose we are given the joint distribution function $F(a, b) = P\{X \leq a, Y \leq b\}$.
- Can we use F to construct a "two-dimensional probability density function"? Precisely, is there a function f such that $P\{(X, Y) \in A\} = \int_A f(x, y) dxdy$ for each (measurable) $A \subset \mathbb{R}^2$?
- Let's try defining $f(x, y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F(x, y)$. Does that work?
- Suppose first that $A = \{(x, y) : x \leq a, \leq b\}$. By definition of F, fundamental theorem of calculus, fact that $F(a, b)$ vanishes as either a or b tends to $-\infty$, we indeed find $\int_{-\infty}^{b}\int_{-\infty}^{a}\frac{\partial}{\partial x}\frac{\partial}{\partial y}F(x, y)dxdy = \int_{-\infty}^{b}\frac{\partial}{\partial y}F(a, y)dy = F(a, b)$.
- From this, we can show that it works for strips, rectangles, general open sets, etc.

Outline

Distributions of functions of random variables

Joint distributions

Independent random variables

Examples

Outline

Independent random variables

- We say X and Y are independent if for any two (measurable) sets A and B of real numbers we have

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}.$$

- Intuition: knowing something about X gives me no information about Y, and vice versa.
- When X and Y are discrete random variables, they are independent if $P\{X = x, Y = y\} = P\{X = x\}P\{Y = y\}$ for all x and y for which $P\{X = x\}$ and $P\{Y = y\}$ are non-zero.
- What is the analog of this statement when X and Y are continuous?
- When X and Y are continuous, they are independent if $f(x, y) = f_X(x)f_Y(y)$.

Sample problem: independent normal random variables

- Suppose that X and Y are independent normal random variables with mean zero and variance one.
- What is the probability that (X, Y) lies in the unit circle? That is, what is $P\{X^2 + Y^2 \leq 1\}$?
- First, any guesses?
- Probability X is within one standard deviation of its mean is about .68. So $(.68)^2$ is an upper bound.
- $f(x,y) = f_X(x) f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-r^2/2}$
- Using polar coordinates, we want $\int_0^1 (2\pi r) \frac{1}{2\pi} e^{-r^2/2} dr = -e^{-r^2/2} \big|_0^1 = 1 - e^{-1/2} \approx .39.$

Outline

Distributions of functions of random variables

Joint distributions

Independent random variables

Examples

Outline

Distributions of functions of random variables

Joint distributions

Independent random variables

Examples

Repeated die roll

- Roll a die repeatedly and let X be such that the first even number (the first 2, 4, or 6) appears on the Xth roll.
- Let Y be the the number that appears on the Xth roll.
- Are X and Y independent? What is their joint law?
- If $j \geq 1$, then

$$P\{X = j, Y = 2\} = P\{X = j, Y = 4\}$$

$$= P\{X = j, Y = 6\} = (1/2)^{j-1}(1/6) = (1/2)^j(1/3).$$

- Can we get the marginals from that?

Continuous time variant of repeated die roll

- On a certain hiking trail, it is well known that the lion, tiger, and bear attacks are independent Poisson processes with respective λ values of $.1/\mathrm{hour}$, $.2/\mathrm{hour}$, and $.3/\mathrm{hour}$.
- Let $T \in \mathbb{R}$ be the amount of time until the first animal attacks. Let $A \in \{\mathrm{lion}, \mathrm{tiger}, \mathrm{bear}\}$ be the species of the first attacking animal.
- What is the probability density function for T? How about $E[T]$?
- Are T and A independent?
- Let T_1 be the time until the first attack, T_2 the subsequent time until the second attack, etc., and let $A_1, A_2, \ldots$ be the corresponding species.
- Are all of the T_i and A_i independent of each other? What are their probability distributions?

More lions, tigers, bears

- Lion, tiger, and bear attacks are independent Poisson processes with λ values $.1/\mathrm{hour}$, $.2/\mathrm{hour}$, and $.3/\mathrm{hour}$.
- Distribution of time T_{tiger} till first tiger attack?
- Exponential $\lambda_{\mathrm{tiger}} = .2/\mathrm{hour}$. So $P\{T_{\mathrm{tiger}} > a\} = e^{-.2a}$.
- How about $E[T_{\mathrm{tiger}}]$ and $\mathrm{Var}[T_{\mathrm{tiger}}]$?
- $E[T_{\mathrm{tiger}}] = 1/\lambda_{\mathrm{tiger}} = 5$ hours, $\mathrm{Var}[T_{\mathrm{tiger}}] = 1/\lambda_{\mathrm{tiger}}^2 = 25$ hours squared.
- Time until 5th attack by any animal?
- Γ distribution with $\alpha = 5$ and $\lambda = .6$.
- X, where Xth attack is 5th bear attack?
- Negative binomial with parameters $p = 1/2$ and $n = 5$.
- Can hiker breathe sigh of relief after 5 attack-free hours?

Buffon's needle problem

- Drop a needle of length one on a large sheet of paper (with evenly spaced horizontal lines spaced at all integer heights).
- What's the probability the needle crosses a line?
- Need some assumptions. Let's say vertical position X of lowermost endpoint of needle modulo one is uniform in $[0, 1]$ and independent of angle θ, which is uniform in $[0, \pi]$. Crosses line if and only there is an integer between the numbers X and $X + \sin\theta$, i.e., $X \leq 1 \leq X + \sin\theta$.
- Draw the box $[0, 1] \times [0, \pi]$ on which (X, θ) is uniform. What's the area of the subset where $X \geq 1 - \sin\theta$?

18.440 Probability and Random Variables
Spring 2014