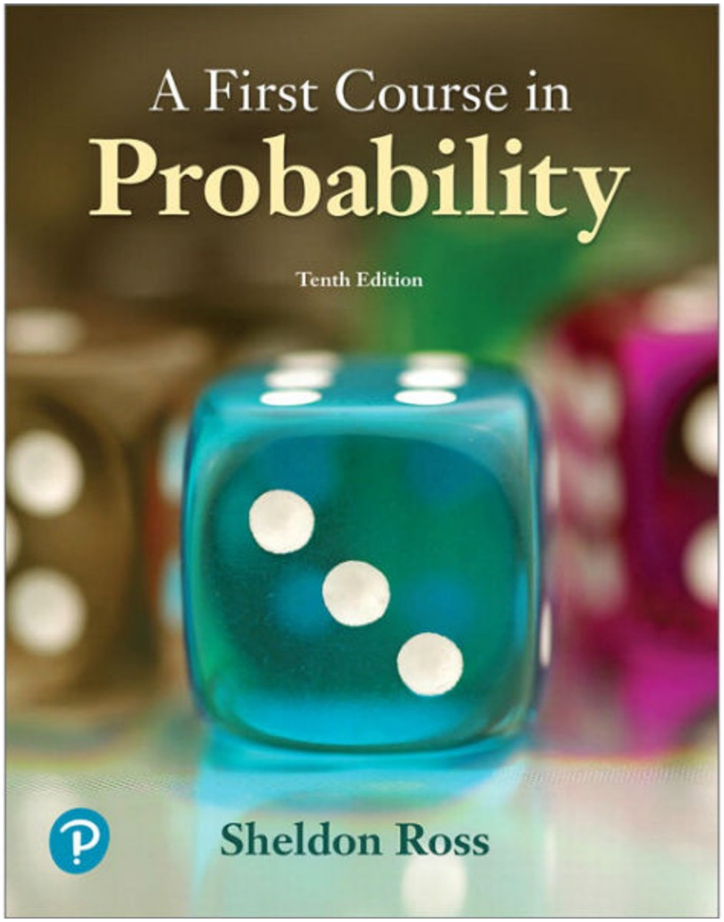
Lecture 11.1

Joint distributions

PURDUE UNIVERSITY | Department of Mathematics



Today's reading: 6.1 (Not on MT2)

Next class: 6.2 (Not on MT2)

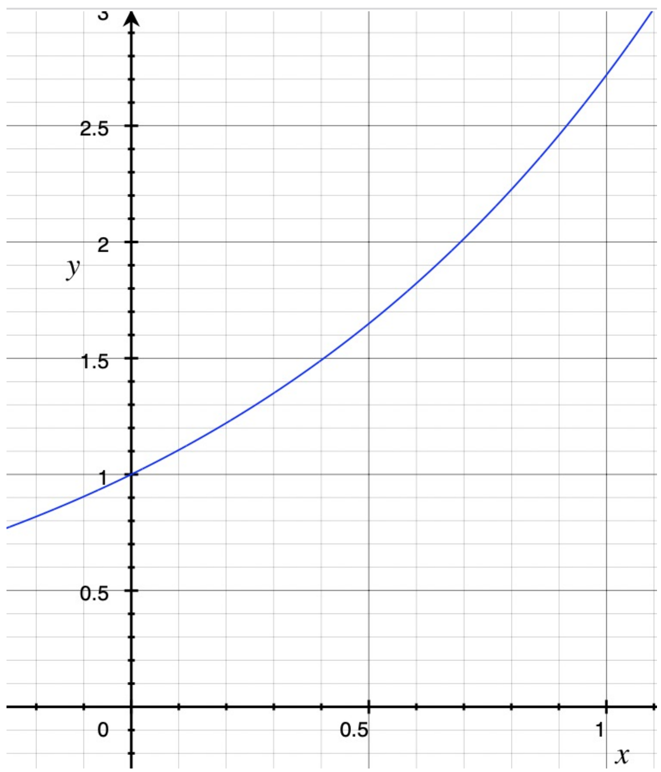
HW9 now available. Due Friday.

I'll do my best to get practice MT2 and studying recommendations to you before class on Wednesday.

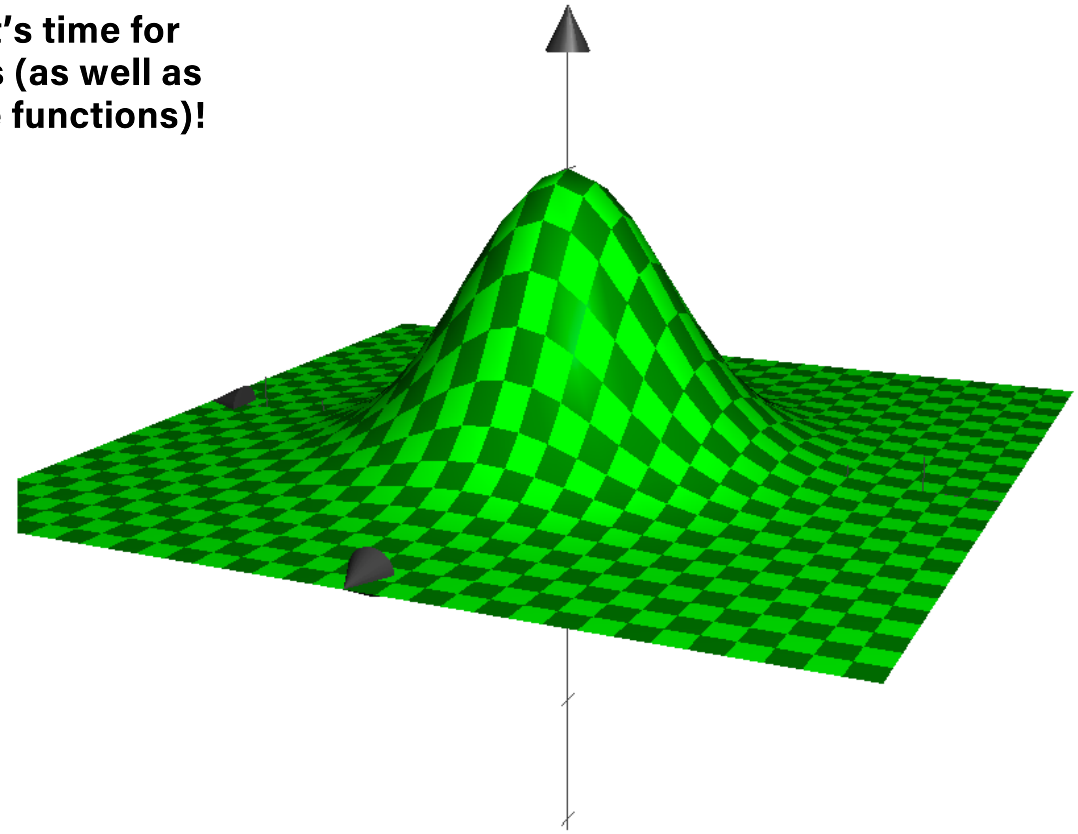
Monday's draft problem

To be presented by Monday's draftee.

If X is uniformly distributed on $(0,1)$, find the density function of $Y = e^X$.



Prepare yourselves: it's time for multivariable calculus (as well as multivariable discrete functions)!



Joint distribution functions

Definition:

Let X and Y be random variables (possibly on different samples spaces S_X, S_Y, with different probability measures P_X, P_Y...). The *joint cumulative distribution function* (*joint CDF*) of X and Y is the function $\mathrm{F}: \mathbb{R}^2 \to \mathbb{R}$ defined as

$$F(a,b) = P\{X \leq a, Y \leq b\}$$

The joint CDF contains most of the information we would ever want to know about both X, Y separately, and how they are related. For example:

1. The CDF of X can be found by taking $F_X(a) = \lim_{b\to+\infty} F(a,b)$.
2. $P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F(a_2,b_2) - F(a_1,b_2) - F(a_2,b_1) + F(a_1,b_1)$. See board for proof by picture.

Joint distribution functions

Definition:

Let X and Y be random variables (possibly on different samples spaces S_X, S_Y, with different probability measures P_X, P_Y...). The *joint cumulative distribution function* (*joint CDF*) of X and Y is the function $\mathrm{F}: \mathbb{R}^2 \to \mathbb{R}$ defined as

$$F(a, b) = P\{X \leq a, Y \leq b\}$$

NOTE WELL: this definition applies equally well in all four possible cases:

1. X is discrete and Y is discrete
2. X is discrete and Y is continuous
3. X is continuous and Y is discrete
4. X is continuous and Y is continuous.

We will be most interested in cases 1 and 4 though.

Jointly discrete random variables

Definition:

Let X and Y be ***discrete*** random variables. The *joint probability mass function* (*joint PMF*) of X and Y is the function $p: \mathbb{R}^2 \to \mathbb{R}$ defined as

$$p(x, y) = P(X = x, Y = y)$$

The joint PMF can be used to recover the individual PMFs. Indeed, let $y_1, y_2, y_3, \ldots$ be the values that Y takes. Then:

$$p_X(x) = P(X = x) = P\left(\bigcup_{i=1}^{\infty} \{X = x, Y = y_i\}\right) = \sum_{i=1}^{\infty} P\{X = x, Y = y_i\} = \sum_{i=1}^{\infty} p(x, y_i)$$

We sometimes call p_X and p_Y the *marginal* PMFs of p.

Example: jointly discrete random variables

Suppose 2 balls are to be selected (without replacement) from an urn that contains 2 red balls, 3 white balls and 4 blue balls. Let X be the number of red balls drawn and let Y be the number of white balls. Let's compute the joint PMF and marginal PMFs of X and Y.

Jointly continuous random variables

Definition:

Let X and Y be random variables. We say they are ***jointly continuous*** if there exists a function $f(x, y)$ such that for every (measurable) event $C \subset \mathbb{R}^2$

$$P\{(X, Y) \in C\} = \iint\limits_{(x,y)\in C} f(x, y)\, dx\, dy$$

We call f the *joint probability density function* (*joint PDF*) of X and Y.
In particular,

$$F(a, b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x, y)\, dx\, dy$$

so by the Fundamental Theorem of Calculus,

$$f(a, b) = \frac{\partial^2}{\partial a\, \partial b} \mathrm{F(a, b)}$$

Jointly continuous random variables

We recover marginals in the continuous case analogously to how we did it in the discrete case:

Let X and Y be jointly continuous random variables. Then

$$P\{X \in A\} = P\{X \in A, -\infty < Y < +\infty\} = \int_A \int_{-\infty}^{+\infty} f(x,y)\, dy\, dx = \int_A f_X(x)\, dx$$

where

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\, dy$$

is the PDF of X.

Let's do examples 1f and 1e now.

SEVERAL jointly random variables

We can generalize all the above from two random variables to several.

Let $X_1, X_2, \dots, X_n$ be several random variables. Their *joint CDF* is defined by

$$F(a_1, a_2, \dots, a_n) = P\{X_1 \le a_1, X_2 \le a_2 \dots, X_n \le a_n\}.$$

We say they are *jointly continuous* if there exists a function $f(x_1, x_2, \dots, x_n)$ such that

$$P\{(X_1, X_2, \dots, X_n) \in C\} = \iint \cdots \int_{(x_1, x_2, \dots, x_n) \in C} f(x_1, x_2, \dots, x_n)\, dx_1 \cdots dx_n$$

We recover the marginal for, say, X_1 by "integrating out" all the other variables:

$$P(X_1 \in A_1) = \int_{A_1} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n)\, dx_2 \cdots dx_{n-1}\, dx_n$$

Wednesday's draft problem

To be presented by Wednesday's draftee.

Suppose X and Y are jointly continuously distributed with joint PDF

$$f(x,y) = \begin{cases} e^{-(x+y)}, & x > 0, y > 0 \\ 0, & \text{else.} \end{cases}$$

Find $P\{X < Y\}$ and $P\{X < a\}$.