Stat 110

Unit 7: Joint Distributions

Ch. 7 in the text

Unit 7 Outline

- Joint, Marginal, and Conditional
  - Discrete, Continuous, & Hybrid
- 2D LOTUS
- Covariance and Correlation
- Multinomial Distribution
- Multivariate Normal Distribution

Joint CDF and PMF

- The joint CDF of r.v.s *X* and *Y* is the function $F_{X,Y}$ given by:

$$F_{X,Y}(x,y) = P(X \le x, Y \le y)$$

- Just like the CDF for a single discrete r.v., the joint CDF is “jumpy”, so usually we consider the joint PMF:
- The joint PMF of two discrete r.v.s *X* and *Y* is the function $F_{X,Y}$ given by:

$$f_{X,Y}(x,y) = P(X = x, Y = y)$$

Joint PMF is a PMF

- The joint PMF must be non-negative and must sum to 1 (just like univariate PMFs):

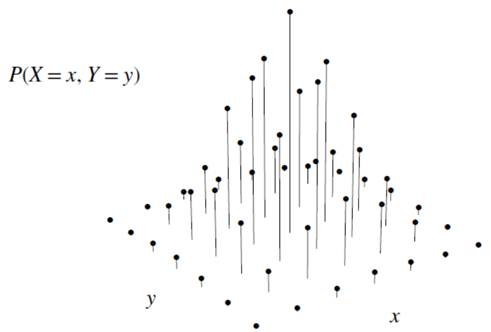
$$\sum_x \sum_y P(X = x, Y = y) = 1$$

- The probability of an event $(X,Y) \in A$ is just the sum of the joint PMF over *A*:

$$P((X,Y) \in A) = \sum \sum_{(x,y)\in A} P(X = x, Y = y)$$

Plot of a Joint PMF

- Joint PMF's can be visualized 3-dimensionally:



Marginal PMF

- For discrete r.v.s X and Y, the ***marginal PMF*** of X is:

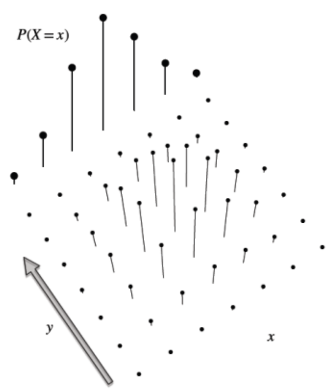
$$P(X = x) = \sum_y P(X = x, Y = y)$$

- So the $P(X = x)$ is the sum over all possible values of Y in the joint PMF where $X = x$.
- We can also define the marginal CDF of X from the joint CDF (much more clumsy):

$$F_X(x) = P(X \le x) = \lim_{y \to \infty} P(X \le x, Y \le y) = \lim_{y \to \infty} F_{X,Y}(x, y)$$

- The marginal PMF can be depicted graphically…

Plot of a marginal PMF vs. joint PMF



Conditional PMF

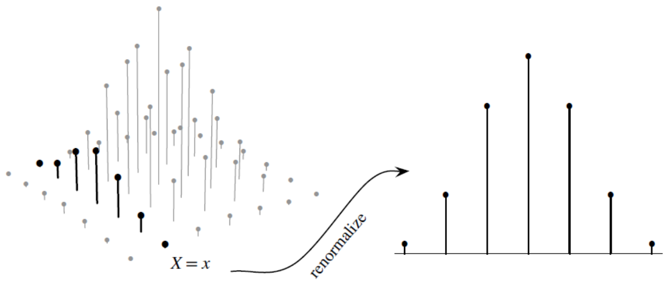
- For discrete r.v.s X and Y, the ***conditional PMF*** of Y given $X{=}x$ is:

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

- What is this a function of? What is random? What is a given?
- So you can think of it as the distribution of Y within the space of X being a specific value, x.
- Bayes rule and the LOTP applies to conditional distributions as well…

$$P(Y = y \mid X = x) = \frac{P(X = x \mid Y = y)P(Y = y)}{\sum_y P(X = x, Y = y)}$$

Plot of a conditional PMF



Joint, Marginal, Conditional: an Example

- The simplest case of a discrete joint distribution is where X and Y are both Bernoulli r.v.s.
- A 2-by-2 contingency table can be created in this setting. Note: this is why it is called *marginal distribution.*
- Let's say that X is the indicator that a patient is obese, and Y is the indicator the patient has diabetes.

	Y = 1	Y = 0
X = 1	10/100	20/100
X = 0	5/100	65/100

- What are the joint, marginal, and conditional distributions of Y conditional on $X = 1$? Conditional on $X = 0$?
- Are X and Y independent?

Independence of Discrete r.v.s

- Two r.v.s X and Y are independent if for all x and y:

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

- If r.v.s X and Y are discrete, this is equivalent to the condition:

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

- Or equivalently:

$$P(Y = y \mid X = x) = P(Y = y)$$

for all y and x such that $P(X = x) > 0$.

Chicken and Egg Problem

- Suppose a chicken lays a random number of eggs, N, with $N \sim \text{Pois}(\lambda)$. Each independently hatches with probability p and fails to hatch with probability $q = 1 - p$. Let X be the number that hatch, and Y be the number that don't hatch ($X + Y = N$). What is the joint PMF of X and Y?
- Key: what should be the conditional PMF of X given N? Of Y given N?
  $(X \mid N = n) \sim \text{Bin}(n,p)$ and $(Y \mid N = n) \sim \text{Bin}[n,(1-p)]$
- What is the conditional PMF of $(X = x, Y = y)$ given N? It's the same as either the PMF of $X = x$ or $Y = y$ (since given N, if I know x, then I automatically know y).
- So that makes life simple:
  $P(X = x, Y = y) = P(X = x \mid N = x + y)P(N = x + y)$.

Chicken and Egg Problem (cont.)

- Thus the joint PMF becomes:

$$P(X=x,Y=y)=P(X=x\mid N=x+y)P(N=x+y)$$
$$=\binom{x+y}{x}p^x(1-p)^y\cdot\frac{e^{-\lambda}\lambda^{x+y}}{(x+y)!}$$
$$=\frac{e^{-\lambda p}(\lambda p)^x}{x!}\cdot\frac{e^{-\lambda(1-p)}(\lambda(1-p))^y}{y!}$$

- So what does this tell us? Since the joint distribution can be factored into a function of x and a separate function of y, then we know that X and Y are independent.
- And, they are marginally Poisson r.v.s!!!

Chicken and Egg Results Summary

- In summary:
- If $X \sim \text{Pois}(\lambda p)$ and $Y \sim \text{Pois}(\lambda(1\text{-}p))$, and X and Y are independent, then $N = X + Y \sim \text{Pois}(\lambda)$ and $(X \mid N = n) \sim \text{Bin}(n,p)$.
  [this is based on the Poisson process of Unit 4].

- And now we know that the converse also holds:

- If $N = X + Y \sim \text{Pois}(\lambda)$ and $(X \mid N = n) \sim \text{Bin}(n,p)$, then $X \sim \text{Pois}(\lambda\, p)$ and $Y \sim \text{Pois}(\lambda(1\text{-}p))$, and X and Y are independent.

Continuous case: Joint PDF

- If X and Y are continuous with joint CDF $F_{X,Y}$ (which needs to be differentiable w.r.t. x and y), their joint PDF is:

$$f_{X,Y}(x,y)=\frac{\partial^2}{\partial x\partial y}F_{X,Y}(x,y)$$

- Joint PDFs must be non-negative and integrate to 1:

$$f_{X,Y}(x,y)\geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy=1$$

- We can get a probability of a two-dimensional region by integrating, for example:

$$P(X<3,1<Y<4)=\int_1^4\int_{-\infty}^3 f_{X,Y}(x,y)dxdy$$

Joint PDF

- We can get a probability of a two-dimensional region by integrating, for example:

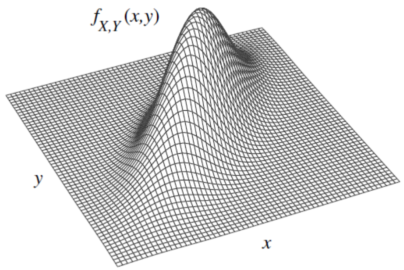
$$P(X<3,1<Y<4)=\int_1^4\int_{-\infty}^3 f_{X,Y}(x,y)dxdy$$

- Or for a general set $A \subseteq \mathbb{R}^2$.

$$P((X,Y)\in A)=\iint_A f_{X,Y}(x,y)dxdy$$

Plot of a Joint PDF

- Joint PDFs can be visualized 3-dimensionally:



Marginal PDF

- For discrete r.v.s X and Y, the ***marginal PDF*** of X is:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

- This is the PDF of X.
- This is easily expandable to more variables:

$$f_{X,W}(x,w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z,W}(x,y,z,w)dydz$$

- Conceptually, this is very straight-forward: just integrate over the unwanted variables to get the [joint] PDF of the wanted variables! But this may not be easy.

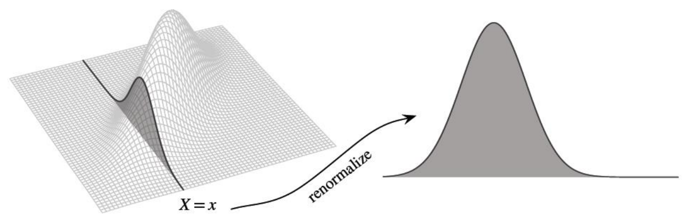
Conditional PDF

- For continuous r.v.s X and Y, the ***conditional PDF*** of Y given X=x is:

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

- This is a function of y for a fixed x. (just like for conditional PMFs and discrete r.v.s)
- The subscripts are there just to be clear that there are 3 separate functions floating around. We could leave them off.
- What does "conditional on x" mean here?

Plot of a conditional PDF



Continuous Bayes' rule and LOTP

- For continuous r.v.s *X* and *Y*, Bayes' rule states:

$$f_{Y|X}(y \mid x) = \frac{f_{X|Y}(x \mid y)\, f_Y(y)}{f_X(x)}$$

- And the Law of Total Probability (LOTP) is:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x \mid y)\, f_Y(y)\,dy$$

- What would happen if we plugged in the other expression for $f_{X,Y}(x,y)$ instead in LOTP?

Independence of Continuous r.v.s

- Two r.v.s *X* and *Y* are independent if for all x and y:

$$F_{X,Y}(x, y) = F_X(x) F_Y(y)$$

- If r.v.s *X* and *Y* are continuous, this is equivalent to the condition:

$$f_{X,Y}(x, y) = f_X(x)\, f_Y(y)$$

- Or equivalently:

$$f_{Y|X}(y \mid x) = f_Y(y)$$

for all *y* and *x* such that $f_X(x) > 0$.

Using PDFs to determine Independence

- Suppose the joint PDF $f_{X,Y}$ for *X* and *Y* factors as:

$$f_{X,Y}(x, y) = g(x) h(y)$$

for all *x* and *y*, where *g* and *h* are non-negative functions.
- Then *X* and *Y* are independent.
- Also, if either *g* or *h* is a valid PDF, then the other is a valid PDF too. And *g* and *h* are the marginal PDFs of *X* and *Y*, respectively.
- Proof: force *h*(*y*) to be a valid PDF (by dividing by a constant *c*), and integrate *y* out. Then you have the marginal PDF for *X*.

Joint PDF example

- Uniform on a square region in the plane:
- Let (*X*,*Y*) be a completely random point in the square {(*x*,*y*): *x*,*y* [0,1]}. What is the joint PDF?

$$f_{X,Y}(x, y) = \begin{cases} 1 & \text{if } x, y \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

- Intuitively, are *X* and *Y* independent?
- What are the marginal PDFs of *X* and *Y*?

Joint PDF example #2

- Uniform on a square region in the plane:
- Let (X,Y) be a completely random point in the unit disk $\{(x,y):\ x^2 + y^2 \le 1\}$. What is the joint PDF?

$$f_{X,Y}(x,y) = \begin{cases} 1/\pi & \text{if } x^2 + y^2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

- Intuitively, are X and Y independent?
  - Key: what is $f_{X|Y}(x|y)$?
- What are the marginal PDFs of X and Y? What are the conditional PDFs?

Joint PDF example #3

- Let $T_1 \sim \text{Expo}(\lambda_1)$ and independently $T_2 \sim \text{Expo}(\lambda_2)$ [just like the #77 and #96 buses to Porter Square].
- Find $P(T_1 < T_2)$.
- Key: what is the joint PDF of (T_1,T_2)? How can we use it to find this probability?

Hybrid Joint Distribution

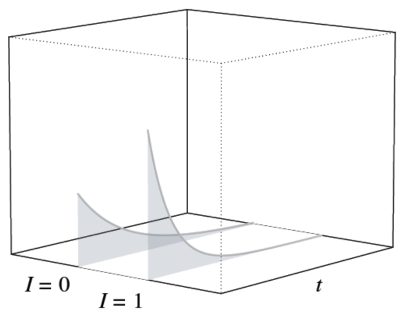
- It is possible to have a joint distribution with 1+ discrete and 1+ continuous r.v.s
- Example: suppose two plants make light bulbs for GE: plant #1's bulbs last $\text{Expo}(\lambda_1)$ and plant #0's last $\text{Expo}(\lambda_0)$. A randomly selected GE bulb has probability p_1 that it was made in plant #1 (and $1-p_1$ it was made in plant #0).
- Let T be how long the bulb lasts, and I be the indicator it was made by plant #1.

  a) Find the CDF and PDF of T.

  b) Does T have the memorylessness property?

  c) Find the conditional distribution of I given $T = t$. What happens to this as $t \to \infty$?

Hybrid Joint Distribution Plot



Hybrid Solutions

a) Find the CDF and PDF of T (for $t > 0$):

$$F_T(t) = P(T \le t) = P(T \le t \mid I = 0)P(I = 0) + P(T \le t \mid I = 1)P(I = 1)$$

$$= (1 - e^{-\lambda_0 t})(1 - p_1) + (1 - e^{-\lambda_1 t})p_1 = 1 - (1 - p_1)e^{-\lambda_0 t} - p_1 e^{-\lambda_1 t}$$

The marginal PDF is the derivative of the above, which is:

$$f_T(t) = \lambda_0(1 - p_1)e^{-\lambda_0 t} + \lambda_1 p_1 e^{-\lambda_1 t}$$

b) Does T have the memorylessness property?
As long as $\lambda_1 \ne \lambda_0$, the PDF does not reduce to the form $\lambda e^{-\lambda t}$ and thus T is not Exponential. This implies that it does not have the memorylessness property (FYI: it is a mixture of 2 Exponentials)

Hybrid Solutions (cont.)

c) Find the conditional distribution of I given $T = t$. What happens to this as $t \to \infty$?

$$P(I = 1 \mid T = t) = \frac{f_{T|I}(t \mid I = 1)P(I = 1)}{f_T(t)}$$

$$= \frac{\lambda_1 e^{-\lambda_1 t} \cdot p_1}{\lambda_0(1 - p_1)e^{-\lambda_0 t} + \lambda_1 p_1 e^{-\lambda_1 t}} = \frac{\lambda_1 p_1}{\lambda_0(1 - p_1)e^{-(\lambda_0 - \lambda_1)t} + \lambda_1 p_1}$$

Thus, the conditional distribution of I given $T = t$ is Bernoulli with this probability of success. If $\lambda_1 > \lambda_0$, this goes to zero as $t \to \infty$ which makes sense intuitively: any bulbs that last a long time is much more likely to come from plant #0 (which has the higher life expectancy).

Unit 7 Outline

- Joint, Marginal, and Conditional
  - Discrete, Continuous, & Hybrid
- 2D LOTUS
- Covariance and Correlation
- Multinomial Distribution
- Multivariate Normal Distribution

2D LOTUS

- If X and Y are discrete r.v.s and g is a function from $\mathbb{R}^2$ to $\mathbb{R}$, then

$$E[g(X,Y)] = \sum_{all\ x} \sum_{all\ y} g(x,y)P(X = x, Y = y)$$

- If X and Y are continuous r.v.s, then:

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

2D LOTUS example

- Let i.i.d X, Y ~Unif(0,1). Find the expected distance between them: $E(|X - Y|)$.

$$E[|X - Y|] = \int_0^1 \int_0^1 |x - y| \, dxdy$$
$$= \int_0^1 \int_y^1 (x - y)dxdy + \int_0^1 \int_0^y (y - x)dxdy$$
$$= 2\int_0^1 \int_y^1 (x - y)dxdy = 1/3$$

Unit 7 Outline

- Joint, Marginal, and Conditional
  - Discrete, Continuous, & Hybrid
- 2D LOTUS
- Covariance and Correlation
- Multinomial Distribution
- Multivariate Normal Distribution

Covariance

- The covariance between r.v.s X and Y is:

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

- This can be expressed equivalently as (using linearity):

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

- Then what is covariance measuring? When will it be large? When will it be small (in magnitude)?
- If two r.v.s are independent, then what should their covariance be?
- Two independent r.v.s will have covariance of zero, and are said to be ***uncorrelated***. Is the converse true?

Properties of Covariance

1. $\text{Cov}(X, X) = \text{Var}(X)$
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
3. $\text{Cov}(X, c) = 0$ for any constant c.
4. $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$ for any constant a.
5. $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$
6. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$

<u>And in General:</u>

$$\text{Var}(X_1 + ... + X_n) = \text{Var}(X_1) + ... + \text{Var}(X_n) + 2\sum_{i<j} \text{Cov}(X_i, X_j)$$

7. What is $\text{Var}(X - Y)$?

Correlation

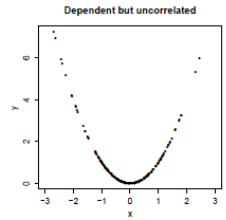
- A similar measure of r.v.s X and Y is the ***correlation*** between them (this is sometimes written as ρ_{XY}):

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

- What are the units on covariance? Units on correlation?
- $-1 \le \mathrm{Corr}(X,Y) \le 1$.
- Proof: Use Var(X + Y) and Var(X – Y).

Independence and Cov & Corr

- As mentioned a few slides earlier, if two r.v.s are independent, then this implies that their covariance is zero (and thus correlation is zero).
- Is the converse true?
- No. What is a Counterexample?
  A "U-shaped" PDF.



- So what is correlation ρ measuring? The linear relationship between X and Y.

Covariance Example

- Let X and Y be i.i.d. Expo(1). Find the correlation between M = max(X,Y) and L = min(X,Y).
- Key: how do M-L and L relate? What are their marginal PDFs? What is their joint PDF?
- M-L and L are independent, and M-L ~ Expo(1) and L ~ Expo(2).
- What is Cov(M, L)?
- Hint: use Cov(M-L+L, L).

Unit 7 Outline

- Joint, Marginal, and Conditional
  - Discrete, Continuous, & Hybrid
- 2D LOTUS
- Covariance and Correlation
- Multinomial Distribution
- Multivariate Normal Distribution

Story of the Multinomial Distribution

- The Multinomial distribution is a generalization of the Binomial. Instead of just 2 possible outcomes, there are now k possible outcomes.
- Imagine you are going to roll a k-sided die n times, with known probability $p_1,\ldots,p_k$ of showing each of the sides, with $p_1+\ldots+p_k = 1$.
- Let X_1 = # of times it shows the first side, X_2 = # of times it shows the second side, etc… so that $X_1 + \ldots + X_n = n$.
- Then the *random vector* $\mathbf{X} = (X_1,\ldots,X_n)$ is said to have the ***Multionomial distribution*** with parameters n and $\mathbf{p}=(p_1,\ldots,p_k)$.
- This is written as $\mathbf{X} \sim \text{Mult}_k(n,\mathbf{p})$.
- What is $\mathbf{X}$'s distribution?

Multinomial Distribution Definition

- If $\mathbf{X} \sim \text{Mult}_k(n,\mathbf{p})$, then the joint PMF of $\mathbf{X}$ is:

$$P(X_1 = x_1,\ldots, X_k = x_k) = \frac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

for $x_1,\ldots x_k$ satisfying $x_1 + \ldots + x_k = n$.

- Note: the n objects in the story of the Multinomial are independent, but the k components are very dependent. Just like in the Binomial: the number of successes and failures are dependent, but the result of each trial is independent from the rest.

More Multinomial Details

- If $\mathbf{X} \sim \text{Mult}_k(n,\mathbf{p})$, then each $X_j \sim \text{Bin}(n,p_j)$ marginally.
- And similarly, we can "lump" categories together, $(X_i + X_j)$ is marginally Binomial, and still have a Multinomial distribution.

- If $\mathbf{X} \sim \text{Mult}_k(n,\mathbf{p})$, then the joint conditional PMF of $\mathbf{X} = (X_2,\ldots,X_n \mid X_1)$ is:

$$(X_2 = x_2,\ldots, X_k = x_k) \mid X_1 = x_1 \sim \text{Mult}_{k-1}(n - x_1, (p_2',\ldots, p_k'))$$

where $p_j' = p_j / (p_2 + \ldots + p_k)$.

Covariance in a Multinomial

- Let $\mathbf{X} \sim \text{Mult}_k(n,\mathbf{p})$. For $i \neq j$, $\text{Cov}(X_i, X_j) = -np_ip_j$.
- Proof:
  Hint: Use $\text{Var}(X_i + X_j)$, the fact that marginally $X_i + X_j \sim$ Binomial and solve for Cov. Use the "lumping" idea.

Unit 7 Outline

- Joint, Marginal, and Conditional
  - Discrete, Continuous, & Hybrid
- 2D LOTUS
- Covariance and Correlation
- Multinomial Distribution
- Multivariate Normal Distribution

Multivariate Normal Distribution

- A random vector $\mathbf{X} = (X_1,\ldots,X_n)$ is said to have ***Multivariate Normal (MVN) distribution*** if every linear combination of the X_j has a Normal distribution. That is we require:

$$t_1 X_1 + \ldots + t_k X_k$$

to have a normal distribution for each choice of $t_1,\ldots, t_k$.
- Note: the degenerate case when all $t_j = 0$ gives a degenerate Normal distribution with variance of zero.
- With the way it is defined, then all marginal distributions of the X_j are Normal.
- A special case is when $k = 2$: the *Bivariate Normal*.

Bivariate Normal PDF

- The joint PDF of a Bivariate Normal (X,Y) with $N(0,1)$ marginal distributions and correlation $\rho \in (-1,1)$ is:

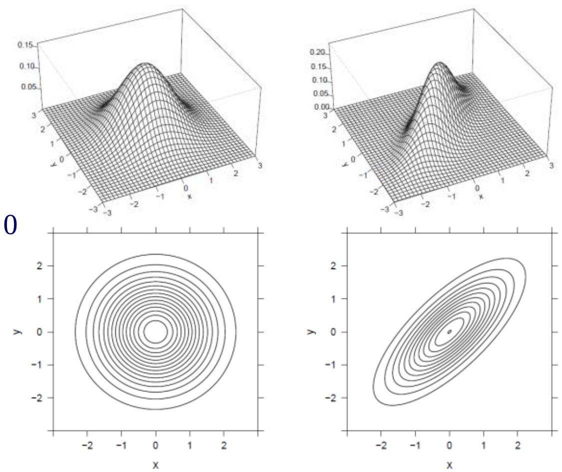
$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right)$$

- Plots of Bivariate Normals are shown on the next slide:

Plot of Bivariate Normals

$\rho = 0$

$\rho = 0.75$



Joint MGFs

- The ***joint moment generating function*** (MGF) of a random vector $\mathbf{X} = (X_1,\ldots,X_n)$ is a function which takes a vector of constants $\mathbf{t} = (t_1,\ldots, t_k)$ and returns:

$$M(\boldsymbol{t}) = E(e^{\boldsymbol{t}'\boldsymbol{X}}) = E(e^{t_1X_1+\ldots+t_kX_k})$$

- This expectation must be finite in a box around the origin in $\mathbb{R}^k$; otherwise we say the joint MGF does not exist.
- Note: we won't worry too much about joint MGFs in this class in general. Just be aware that they exist, and that they fully determine the joint distribution of $\mathbf{X}$.
- They are particularly nice for Multivariate Normals.

Joint MGF of MVNs

- Recall, the MGF of a Normal distribution:

$$E(e^W) = e^{E(W)+Var(W)/2}$$

- Therefore, the joint MGF of a MVN $(X_1,\ldots,X_n)$ is:

$$M(\boldsymbol{t}) = E(e^{t_1X_1+\ldots+t_kX_k})$$
$$= \exp\left[t_1E(X_1)+\ldots+t_kE(X_k)+\mathrm{Var}(t_1X_1+\ldots+t_kX_k)/2\right]$$

Correlation of zero is equivalent to independence for MVNs

- A very important property of Multivariate Normals is that a correlation of zero ($\boldsymbol{\rho} = 0$) means that every pair of components X_i, X_j are independent of one another (and we already know the converse holds for all r.v.s).
- This can be proved based on joint MGFs of Bivariate Normals. The extension to general MVNs is analogous.
- Let (X,Y) be Bivariate Normal with means μ_X and μ_Y, variances σ^2_X and σ^2_Y, and correlation ρ. Find the joint MGF. Then let $\rho = 0$.
- This is equivalent then to the joint MGF of independent $Z \sim N(\mu_X, \sigma^2_X)$ and $W \sim N(\mu_Y, \sigma^2_Y)$. Since joint MGFs determine joint distributions, then we know X and Y are independent.

Independence of sum and difference for 2 Normals

- Let X,Y be i.i.d. $N(0,1)$. Find the joint distribution of $(X+Y, X–Y)$.
- What distribution is $(X+Y)$? What about $(X–Y)$?
- What about their joint distribution?
- What is the Correlation of $(X+Y)$ and $(X–Y)$? First find the Covariance.

Bivariate Normal Generation

- Suppose we can create realizations of i.i.d. r.v.s $X,Y \sim N(0,1)$, but want to generate a Bivariate normal (Z,W) with desired $\mathrm{Corr}(Z,W) = \rho$. How can we construct Z and W from a linear combinations of X and Y?
- Key: let $Z = aX + bY$ and $W = cX + dY$.
- What will be the means of Z and W?
- What are the variances of Z and W in terms of a,b,c,d?
- What is the covariance?
- 4 unknowns and 3 equations, but we only need one solution. So let $b = 0$, and solve:

$$Z = X$$
$$W = \rho X + \left(\sqrt{1-\rho^2}\right)Y$$