CEE518: Functions of Random Variables

$$Y = g(X)$$

$$F_Y(y) = P(Y \le y) = P(g(X) \le y) = \int_{g(X)\le y} f_X(x)\,dx$$

$$f_Y(y) = \frac{d}{dy}\left[F_Y(y)\right]$$

If g(X) is a monotonically increasing or decreasing function,

$$x = g^{-1}(y) = h(y)$$

where X can be seen to be a single-valued function of y.

$$dx = \left|\frac{dh}{dy}\right| dy = \left|h^{'}(y)\right| dy$$

These equations lead to

$$F_Y(y) = \int_{-\infty}^{y} f_X[h(y)]\left|h^{'}(y)\right| dy$$

$$f_Y(y) = f_X[h(y)]\left|h^{'}(y)\right|$$

Example. Assume that $Y = a + bX$, for $a > 0$ and $b > 0$. Derive the probability density of Y in terms of X

$$x = \frac{y-a}{b} = h(y)$$

$$\frac{dh}{dy} = \frac{1}{b}$$

$$f_Y(y) = f_X[h(y)]\left|h'(y)\right|$$

$$= \frac{1}{b} f_X\left(\frac{y-a}{b}\right) f_Y(y) = f_X[h(y)]\left|h'(y)\right| = \frac{1}{b} f_X\left(\frac{y-a}{b}\right)$$

Consider a function of two random variables X_1 and X_2 as

$$Y = g(X_1, X_2)$$

$$F_Y(y) = \iint f_{X_1 X_2}(x_1, x_2)\, dx_1 dx_2$$

$$x_1 = g^{-1}(Y, x_2)$$

$$F_Y(y) = \int_{x_2=-\infty}^{\infty} \int_{y=-\infty}^{y} f_{X_1 X_2}(g^{-1}, x_2) \left|\frac{\partial g^{-1}}{\partial y}\right| dy dx_2$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1 X_2}(g^{-1}, x_2) \left|\frac{\partial g^{-1}}{\partial y}\right| dx_2$$

If X_1 and X_2 are independent,

$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)\, f_{X_2}(x_2)$ and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(y - x_2)\, f_{X_2}(x_2)\, dx_2$$

If

$Y = X_1 + X_2$

$$F_Y(y) = \int_{x_1 + x_2 \le y} \int_{x_1 + x_2 \le y} f_{X_1,X_2}(x_1, x_2) dx_1 dx_2$$

$$F_Y(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{y - x_2} f_{X_1,X_2}(x_1, x_2) dx_1 dx_2$$

Differentiation yields

$$f_Y(y) = \frac{dF_Y}{dy} = \int_{-\infty}^{\infty} f_{X_1,X_2}(y - x_2, x_2) dx_2$$

Example. Find the probability density of the sum of two independent normally distributed random variables.

Let X_1 and X_2 be normal variates with

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}\sigma_{X_1}} e^{-\frac{1}{2}\left(\frac{x_1 - \mu_{X_1}}{\sigma_{X_1}}\right)^2}$$

and

$$f_{X_2}(x_2) = \frac{1}{\sqrt{2\pi}\sigma_{X_2}} e^{-\frac{1}{2}\left(\frac{x_2 - \mu_{X_2}}{\sigma_{X_2}}\right)^2}$$

The substitution of these two into the equation for $f_Y(y)$ yields

$$f_Y(y)=\frac{dF_Y}{dy}=\int f_{X_1X_2}(y-x_2,x_2)dx_2$$

$$=\frac{1}{2\pi\sigma_{X_1}\sigma_{X_2}}\int_{-\infty}^{+\infty}\exp-\frac{1}{2}\left[\left\{\left(\frac{y-x_2-\mu_{X_1}}{\sigma_{X_1}}\right)^2+\left(\frac{x_2-\mu_{X_2}}{\sigma_{X_2}}\right)^2\right\}\right]dX_2$$

$$f_Y(y)=\frac{1}{\sqrt{2\pi\left(\sigma_{X_1}^2+\sigma_{X_2}^2\right)}}\exp\left[-\frac{1}{2}\left(\frac{y-(\mu_{X_1}+\mu_{X_2})}{\sqrt{\sigma_{X_1}^2+\sigma_{X_2}^2}}\right)^2\right]$$

$$\therefore\ \mu_Y=\mu_{X_1}+\mu_{X_2}$$
$$\sigma_Y^2=\sigma_{X_1}^2+\sigma_{X_2}^2$$

This result shows that Y is also a normal variate with

$$\mu_Y=\mu_{X_1}+\mu_{X_2}$$
$$\sigma_Y^2=\sigma_{X_1}^2+\sigma_{X_2}^2$$

For the case of a linear sum of *n* random variables: If

$$Y=\sum_{i=1}^{n} a_iX_i$$

where a_i are constants and X_i are independent normal variates $N(\mu_{X_i},\sigma_{X_i})$ then Y is also a normal variate $N(\mu_Y,\sigma_Y)$ with

$$\mu_Y=\sum_{i=1}^{n} a_i\mu_{X_i}$$
$$\sigma_Y^2=\sum_{i=1}^{n} a_i^2\sigma_{X_i}^2$$

Mean and variance of a linear function

If

$$Y = a_1X_1 + a_2X_2 + \quad + a_nX_n$$

where the a_i are constants and the X_i are random variables with means μ_i and standard deviations σ_i. The mean value of Y is by definition

$$\mu_Y = E[Y] = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = a_1\mu_1 + \quad + a_n\mu_n$$

and the variance of Y is given by

$$\mathrm{var}[Y] = \sigma_Y^2 = E\left[(Y-\mu_Y)^2\right] = \sum_{i=1}^{n} a_i^2 \mathrm{var}[X_i] + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \rho_{ij}\sigma_i\sigma_j$$

Mean and sum of two random variables

If $Y = X_1 \pm X_2$ then

$$\mu_Y = \mu_{X_1} \pm \mu_{X_2}$$

and

$$\sigma_Y = \begin{Bmatrix} \sqrt{\sigma_{X_1}^2 + \sigma_{X_2}^2} & \text{if } X_1 \text{ and } X_2 \text{ are uncorrelated} \\ \sqrt{\sigma_{X_1}^2 + \sigma_{X_2}^2 \pm 2\rho\sigma_{X_1}\sigma_{X_2}} & \text{if } X_1 \text{ and } X_2 \text{ are correlated} \end{Bmatrix}$$

where ρ is the correlation coefficient between X_1 and X_2.

If

$Y = X_1 X_2$

then

$$\mu_Y = E[X_1 X_2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{X_1,X_2}(x_1, x_2)\, dx_1 dx_2$$

and if X_1 and X_2 are independent,

$$f_{X_1,X_2} = f_{X_1}(x_1) f_{X_2}(x_2) f_{X_1,X_2} = f_{X_1}(x_1) f_{X_2}(x_2)$$

and

$$\mu_Y = \mu_{X_1} \mu_{X_2}$$

$$\sigma_Y^2 = E\left[(Y - \mu_Y)^2\right] = E\left[X_1^2 X_2^2\right] + \mu_{X_1}^2 \mu_{X_2}^2 - 2\mu_{X_1}\mu_{X_2} E[X_1 X_2]$$

If X_1 and X_2 are independent, then

$$E\left[X_1^2 X_2^2\right] = \mu_{X_1}^2 \mu_{X_2}^2.$$

For normally distributed variables,

$$E\left[X^2\right] = \mu_X^2 + \sigma_X^2$$

and

$$\sigma_Y = \sqrt{\mu_{X_1}^2 \sigma_{X_2}^2 + \mu_{X_2}^2 \sigma_{X_1}^2 + \sigma_{X_1}^2 \sigma_{X_2}^2}$$

If X_1 and X_2 are correlated with a correlation coefficient of ρ, the standard deviation of Y is

$$\sigma_Y = \mu_{X_1}\mu_{X_2} \sqrt{\left(\frac{\sigma_{X_1}^2}{\mu_{X_1}^2} + \frac{\sigma_{X_1}^2}{\mu_{X_2}^2} + \frac{\sigma_{X_1}^2 \sigma_{X_2}^2}{\mu_{X_1}^2 \mu_{X_2}^2}\right)\left(1 + \rho^2\right)}$$

If

$$Y = \frac{X_1}{X_2}$$

the mean and variance can be obtained for normal variates X_1 and X_2

$$\mu_Y = \frac{\mu_{X_1}}{\mu_{X_2}}$$

$$\sigma_Y = \begin{cases} \dfrac{1}{\mu_{X_2}} \sqrt{\dfrac{\mu_{X_1}^2 \sigma_{X_2}^2 + \mu_{X_2}^2 \sigma_{X_1}^2}{\mu_{X_2}^2 + \sigma_{X_2}^2}} & \text{if } X_1 \text{ and } X_2 \text{ are uncorrelated} \\ \dfrac{\mu_{X_1}^2}{\mu_{X_2}^2} \sqrt{\dfrac{\sigma_{X_1}^2}{\mu_{X_1}^2} + \dfrac{\sigma_{X_2}^2}{\mu_{X_2}^2} - 2\rho \dfrac{\sigma_{X_1}\sigma_{X_2}}{\mu_{X_1}\mu_{X_2}}} & \text{if } X_1 \text{ and } X_2 \text{ are correlated} \end{cases}$$

For a nonlinear function of several random variables

$$Y = g(X_1, X_2, \ldots, X_n)$$

the mean and variance are

$$\mu_Y = E[Y] = \int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_n$$

$$\sigma_Y^2 = E\left[(Y - \mu_Y)^2\right] = \int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} \left[g(x_1, x_2, \ldots, x_n) - \mu_Y\right]^2 f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_n$$

Taylor Series Approximation

$$Y = g(\mu_1, \mu_2, \ldots, \mu_n) + \sum_{i=1}^{n} \frac{\partial g}{\partial X_i}\bigg|_{\mu_1,\mu_2,\ldots,\mu_n} (X_i - \mu_i)$$

$$+ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 g}{\partial X_i \partial X_j}\bigg|_{\mu_1,\mu_2,\ldots,\mu_n} (X_i - \mu_i)(X_j - \mu_j) +$$

where the partial derivatives are evaluated at the mean values of X_i as

$$\frac{\partial g}{\partial X_i}\bigg|_{\mu_1,\mu_2,\ldots,\mu_n} = \frac{\partial g}{\partial X_i}(X_1 = \mu_1, \ldots, \quad X_n = \mu_n)$$

By neglecting higher order derivatives

$$Y \approx g(\mu_1, \mu_2, \ldots, \mu_n) + \sum_{i=1}^{n} \frac{\partial g}{\partial X_i}\bigg|_{\mu_1,\mu_2,\ldots,\mu_n} (X_i - \mu_i)$$

The mean value of Y becomes

$$\mu_Y = E[Y] \approx g(\mu_1, \mu_2, \ldots, \mu_n)$$

The variance of Y becomes

$$\sigma_Y^2 \approx \sum_{i=1}^{n} a_i^2 \operatorname{var}[X_i] + \sum_{i=1}^{n} \sum_{j=1,\, i \neq j}^{n} a_i a_j \operatorname{cov}[X_i, X_j]$$

where

$$a_i = \frac{\partial g}{\partial X_i}\bigg|_{\mu_1,\mu_2,\ldots,\mu_n}$$

and

$$a_j = \frac{\partial g}{\partial X_j}\bigg|_{\mu_1,\mu_2,\ldots,\mu_n}$$

If the variables X_i are statistically independent, then

$$\sigma_Y^2 \approx \sum_{i=1}^{n} a_i^2 \sigma_{X_i}^2$$

This procedure is known as the partial derivative rule.

Example problem: (Kapur & Lamberson, '77)

To compute the tensile stresses induced in a circular bar of diameter D, we need to know the cross-section $A = \frac{\pi}{4} D^2$. Suppose the diameter has a normal distribution with a mean μ and standard deviation σ. We wish to find the density function for the random variable A.

$$A = f(D) = \frac{\pi}{4} D^2. \qquad D = f^{-1}(A) = \pm\sqrt{\frac{4A}{\pi}}$$

Taking the derivative yields

$$\left|\frac{dD}{dA}\right| = \sqrt{\frac{1}{A\pi}}$$

The inverse function is double-valued so

$$h(A) = \frac{1}{A\pi} g\left(\sqrt{\frac{4A}{\pi}}\right) + g\left(-\sqrt{\frac{4A}{\pi}}\right)$$

If g is given to be a normal density function with mean μ and standard deviation σ, then

$$h(A) = \frac{\left(1/A\pi\right)}{\sigma\sqrt{2\pi}}\left\{\exp\left[-\frac{\left(\sqrt{4A/\pi} - \mu\right)^2}{2\sigma^2}\right] + \exp\left[-\frac{\left(-\sqrt{4A/\pi} - \mu\right)^2}{2\sigma^2}\right]\right\}$$

Example Problem

•The load P acting on a bar in tension has a mean value $\mu_P = 10{,}000\ N$ and a standard deviation $\sigma_P = 1{,}000\ N$. The mean value of the cross-sectional area A is $\mu_A = 5.0\ cm^2$ and the standard deviation of A is $\sigma_A = 0.4\,cm$. Find the mean and standard deviation of the tensile stress s on the bar. Use the Taylor Series Approximation.

•The moment equation for a fixed-fixed beam with a concentrated load P and uniformly loaded w is

$$M = \frac{PL}{8} + \frac{wL^2}{12}$$

If the means and standard deviations are $\overline{P}, \overline{w}, \sigma_P$ and σ_w and $Cov(P,w) = 0$, then find $\overline{M}$ and σ_M.

Let $\Delta = \frac{PL^3}{192EI} + \frac{wL^4}{384EI}$ Find $\overline{\Delta}, \sigma_\Delta$ and $Cov(\Delta, M)$.

•A connecting rod of length L and diameter d is subjected to an axial compressive load, P. The Euler buckling load is

$$P_c = \frac{\pi^2 EI}{L^2}$$

The parameters are given as

L = N(20,0.5)inches

$d = N(\overline{d}, 0.1\overline{d})$

P = N(2000,200)lbs

$E = N(30\text{x}10^6, 3\text{x}10^6)psi$

Design the rod to achieve a reliability of 0.99 against buckling. Hint: Use

$$I = \frac{\pi d^4}{64}$$

Let Strength R - Stress Q = g(R,Q) = Y

If both strength and stress are Normally distributed,

the reliability R_0 can be written as

$$R_0 = \frac{1}{\sqrt{2\pi}} \int_{s_1}^{\infty} \exp[-\frac{1}{2}s^2]ds$$

where

$$s_1 = \left(\frac{\mu_R - \mu_Q}{\sigma_R^2 + \sigma_Q^2} \right)$$

This allows us to use the standard Normal tables.

$$R_0 = 0.99 \text{ corresponds to } s_1 = -2.32635 = -\left\{ \frac{\overline{P}_C - \overline{P}}{\sqrt{\sigma_{P_c}^2 + \sigma_P^2}} \right\}$$

•Given

$$a\left({cm}/{s^2} \right) = 1230e^{0.8M}(R+2.5)^{-2}$$

where a = maximum ground acceleration at a distance R kilometers from an earthquake of Richter magnitude M.

Assume $M = 6$ and consider it deterministic.

Consider R = r.v. with mean 25 km and COV of 10%.

Use a <u>*linear*</u> statistical analysis to find a, σ_a *and COV*.