Conditional expectation and prediction

Conditional frequency functions and pdfs have properties of ordinary frequency and density functions. Hence, associated with a conditional distribution is a conditional mean.

Y and *X* are discrete random variables, the conditional frequency function of *Y* given *x* is $p_{Y|X}(y|x)$.

Conditional expectation *of Y given X=x is*

$$E(Y|X=x)=\sum_{y} yp_{Y|X}(y|x)$$

Continuous case:

$$E(Y|X=x)=\int yf_{Y|X}(y|x)\,dy$$

Conditional expectation of a function:

$$E[h(Y)|X=x]=\int h(y)\,f_{Y|X}(y|x)\,dy$$

Consider a Poisson process on [0, 1] with mean λ, and let N be the # of points in [0, 1]. For $p < 1$, let X be the number of points in [0, p]. Find the conditional distribution and conditional mean of X given $N = n$.

Make a guess!

Consider a Poisson process on [0, 1] with mean λ, and let N be the # of points in [0, 1]. For $p < 1$, let X be the number of points in [0, p]. Find the conditional distribution and conditional mean of X given $N = n$.

We first find the joint distribution: $P(X = x, N = n)$, which is the probability of x events in $[0, p]$ and $n-x$ events in $[p, 1]$.

From the assumption of a Poisson process, the counts in the two intervals are independent Poisson random variables with parameters $p\lambda$ and $(1-p)\lambda$ (**why?**), so

$$p_{XN}(x, n) = \frac{(p\lambda)^x e^{-p\lambda}}{x!} \frac{[(1-p)\lambda]^{n-x} e^{-(1-p)\lambda}}{(n-x)!}$$

N has Poisson marginal distribution, so the conditional frequency function of X is

$$p_{X|N}(x|n) = \frac{p_{XN}(x, n)}{p_N(n)}$$

$$= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

Binomial distribution,
Conditional expectation is np.

Conditional expectation of Y given $X=x$ is a function of X, and hence also a random variable, $E(Y|X)$.

In the last example, E(X|N=n)=np, and E(X|N)=Np is a function of N, a random variable that generally has an expectation

$$E[E(Y|X)]$$

Taken w.r.t. the distribution of X

Law of total expectation

The average (expected) value of Y can be found by first conditioning on X, finding $E(Y|X)$, and then averaging this quantity with respect to X.

$$E(Y) = E[E(Y|X)].$$

In other words, the expectation of a random variable Y can be calculated by weighting the conditional expectations appropriately and summing or integrating.

$$E(Y) = E[E(Y|X)].$$

Proof. RHS = $\sum_x E(Y|X=x)p_X(x)$

$$E(Y|X=x) = \sum_y y p_Y(y|x)$$

Interchanging the order of summation,

$$\sum_x E(Y|X=x)p_X(x) = \sum_y y \sum_x p_{Y|X}(y|x)p_X(x)$$

Law of total probability, $p_Y(y)$

Finally, RHS $= \sum_y y p_Y(y) = E(Y)$ = LHS.

Example.
In a system, a component and a backup unit both have mean lifetimes equal to μ. If the component fails, the system automatically substitutes the backup unit, but there is probability p that something will go wrong and it will fail to do so. What is the expected total lifetime?

Let T be the total lifetime, let $X = 1$ if the substitution of the backup takes place successfully, and $X = 0$ if it does not.

$$E(T|X=1) = 2\mu$$
$$E(T|X=0) = \mu$$

$$\begin{aligned} &E(T) \\ &\quad = E(T|X=1)P(X=1) + E(T|X=0)P(X=0) \\ &\quad = \mu(2-p) \end{aligned}$$

Example. ***Random sums.***

1. 保险公司在一年内收到N笔索赔，每笔数量为$X_1, X_2, \ldots, X_n$,总赔款为T。
2. 商场来了N个客户，每位消费了$X_1, X_2, \ldots, X_n$,总销售额度为T。
3. 3月6日，N个人在东活排成一队，报名参加女生节校园活动，每人用时$X_1, X_2, \ldots, X_n$,总排队时间为T。

$E(T)$?

Make a guess!

Example. ***Random sums.***

1. 保险公司在一年内收到N笔索赔，每笔数量为$X_1, X_2, \ldots, X_n$,总赔款为T。
2. 商场来了N个客户，每位消费了$X_1, X_2, \ldots, X_n$,总销售额度为T。
3. 3月6日，N个人在东活排成一队，报名参加女生节校园活动，每人用时$X_1, X_2, \ldots, X_n$,总排队时间为T。

$$T = \sum_{i=1}^{N} X_i \qquad\qquad E(T) = E[E(T|N)]$$

Since $E(T|N = n) = nE(X)$, $E(T|N) = NE(X)$ and thus

$$E(T) = E[NE(X)] = E(N)E(X)$$

What about the variance? (We do not prove it.)

$$\mathrm{Var}(Y) = \mathrm{Var}[E(Y|X)] + E[\mathrm{Var}(Y|X)]$$

Example. ***Random sums*** again.
Additional assumptions: X_i are independent random variables with the same mean, $E(X)$, and the same variance, $\mathrm{Var}(X)$.

$$\mathrm{Var}(T) = E[\mathrm{Var}(T|N)] + \mathrm{Var}[E(T|N)]$$

Because $E(T|N) = NE(X)$,

$$\mathrm{Var}[E(T|N)] = [E(X)]^2\mathrm{Var}(N)$$

Also, since $\mathrm{Var}(T|N = n) = \mathrm{Var}(\sum_{i=1}^{n} X_i) = n\,\mathrm{Var}(X)$,

$$\mathrm{Var}(T|N) = N\,\mathrm{Var}(X)$$

$$E[\mathrm{Var}(T|N)] = E(N)\mathrm{Var}(X)$$

$$\mathrm{Var}(T) = [E(X)]^2\mathrm{Var}(N) + E(N)\mathrm{Var}(X)$$

Example. ***Random sums*** again.

$$\mathrm{Var}(T) = [E(X)]^2 \mathrm{Var}(N) + E(N)\mathrm{Var}(X)$$

Suppose that # of insurance claims in a certain time period (a Poisson random variable) has expected value 900 and standard deviation 30. Suppose that the average claim value is $1000 and the standard deviation is $500.

Then the expected value of the total, T, of the claims is

$$E(T) = 900 \cdot 1000 = \$900{,}000$$

and the variance of T is

$$\mathrm{Var}(T) = 1000^2 \cdot 30^2 + 900 \cdot 500^2 = 1.125 \cdot 10^9$$

Standard deviation is $33541.

If total # is not variable, fixed at N=900, $\mathrm{Var}(T)=E(N)\ \mathrm{Var}(X)$, standard deviation is $15,000, much smaller.

Prediction

Predicting the value of one random variable from another

Examples.

1. Predict the age of a fish through measuring its length. Lengths and ages are joint random variables.
2. In forestry, estimate the volume of a tree from its diameter. Diameter and volume are joint random variables.
3. Predict your fate from your appearance. (??!!)

Examples.

1. Predict the age of a fish through measuring its length. Lengths and ages are joint random variables.
2. In forestry, estimate the volume of a tree from its diameter. Diameter and volume are joint random variables.
3. Predict your fate from your appearance. (??!!)

Trivial case: predicting *Y* by means of a constant value *c*.

Need some measure of the prediction effectiveness, widely used mean squared error should be minimized:

$$\text{MSE} = E[(Y - c)^2]$$

$$\begin{aligned} E[(Y - c)^2] &= \text{Var}(Y - c) + [E(Y - c)]^2 \\ &= \text{Var}(Y) + (\mu - c)^2 \end{aligned}$$

-- Should use $c=\mu=E(Y)$.

Predicting *Y* by some function *h*(*X*).

Minimize MSE=$E\{[Y-h(X)]\}$.

$$E\{[Y-h(X)]^2\} = E(E\{[Y-h(X)]^2|X\})$$

For every x, the inner expectation is minimized by setting $h(x)$ equal to the constant $E(Y \mid X=x)$, according to the preceding trivial case.

$$h(X) = E(Y|X)$$

Unfortunately, this optimal prediction scheme depends on knowing the joint distribution of Y and X to find $E(Y|X)$, which is often N/A.

Predicting *Y* by some function *h*(*X*).

Minimize MSE=$E\{[Y-h(X)]\}$.

$$E\{[Y-h(X)]^2\} = E(E\{[Y-h(X)]^2|X\})$$

For every x, the inner expectation is minimized by setting $h(x)$ equal to the constant $E(Y \mid X=x)$, according to the preceding trivial case.

$$h(X) = E(Y|X)$$

Unfortunately, this optimal prediction scheme depends on knowing the joint distribution of Y and X to find $E(Y|X)$, which is often N/A.

Instead of trying to find the best function h among all function, we optimize our *linear* predictor (最优线性预测元) $h(x) = \alpha + \beta x$.

$$\begin{aligned} E[(Y-\alpha-\beta X)^2] &= \mathrm{Var}(Y-\alpha-\beta X) + [E(Y-\alpha-\beta X)]^2 \\ &= \underbrace{\mathrm{Var}(Y-\beta X)}_{\text{(no } \alpha\text{)}} + \underbrace{[E(Y-\alpha-\beta X)]^2}_{\text{(zero if } \alpha = \mu_Y - \beta\mu_X\text{)}} \end{aligned}$$

$$E[(Y-\alpha-\beta X)^2] = \mathrm{Var}(Y-\alpha-\beta X) + [E(Y-\alpha-\beta X)]^2$$
$$= \mathrm{Var}(Y-\beta X) + [E(Y-\alpha-\beta X)]^2$$

(no α) (zero if $\alpha = \mu_Y - \beta\mu_X$)

$$\mathrm{Var}(Y-\beta X) = \sigma_Y^2 + \beta^2\sigma_X^2 - 2\beta\sigma_{XY}$$

Minimum of the quadratic function of β is found by setting the derivative w.r.t. β equal to zero

$$\beta = \frac{\sigma_{XY}}{\sigma_X^2} = \rho\frac{\sigma_Y}{\sigma_X}$$

ρ=correlation coefficient

The minimum MSE predictor is

$$\hat{Y} = \alpha + \beta X = \mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}(X-\mu_X)$$

The mean squared prediction error is

$$\mathrm{Var}(Y-\beta X) = \sigma_Y^2 + \frac{\sigma_{XY}^2}{\sigma_X^4}\sigma_X^2 - 2\frac{\sigma_{XY}}{\sigma_X^2}\sigma_{XY}$$
$$= \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} = \sigma_Y^2 - \rho^2\sigma_Y^2 = \sigma_Y^2(1-\rho^2)$$

Notes:

1. Optimal linear predictor depends on joint distribution of X, Y only through their means, variances, and covariance, unlike general optimal predictor $E(Y|X)$.
2. Mean squared prediction error depends only on σ_Y and ρ, small if ρ is about ± 1.
3. For the bivariate normal distribution, direct calculation shows exactly the same form as the optimal linear predictor

$$E(Y|X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)$$

Notes:

1. Optimal linear predictor depends on joint distribution of X, Y only through their means, variances, and covariance, unlike general optimal predictor $E(Y|X)$.
2. Mean squared prediction error depends only on σ_Y and ρ, small if ρ is about ± 1.
3. For the bivariate normal distribution, direct calculation shows exactly the same form as the optimal linear predictor

$$E(Y|X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)$$

Example. Two exams are given in a course. The scores of a student on the mid-term and final exams, X and Y, are jointly distributed. Suppose that the exams are scaled to have the same means $\mu = \mu_X = \mu_Y$ and standard deviations $\sigma = \sigma_X = \sigma_Y$. Then, the correlation $\rho = \sigma_{XY}/\sigma^2$ and the best linear predictor $\hat{Y} = \mu + \rho(X - \mu)$.

By the equation

$$\hat{Y} - \mu = \rho(X - \mu)$$

we predict that student's score on the final exam to differ from the overall mean μ by less than did the score on the mid-term. In case of positive correlation:

-- Encouraging for students below average, bad news for those above average

This phenomenon is often referred to as regression to the mean.

The moment-generating functions

Moment-generating functions (mgf, 矩母函数、矩生成函数) are very useful tools that can dramatically simplify certain calculations.

$$M(t) = E(e^{tX})$$

Discrete case:

$$M(t) = \sum_x e^{tx} p(x)$$

Continuous case:

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

PROPERTY **A**

If the moment-generating function exists for t in an open interval containing zero, it uniquely determines the probability distribution. ■

If two random variables have the same mgf in an open interval containing zero, they have the same distribution.

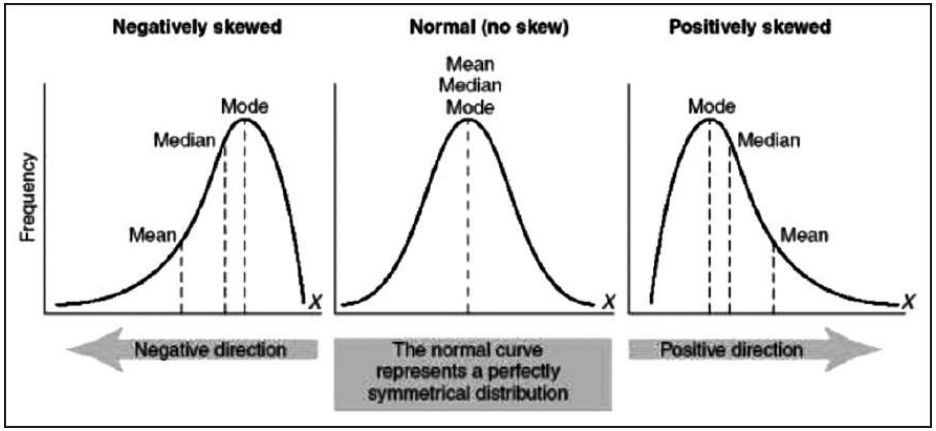
For some problems, we can find the mgf and then deduce the unique probability distribution corresponding to it.

The ***r*th moment** of a random variable, if exists, is $E(X^r)$.

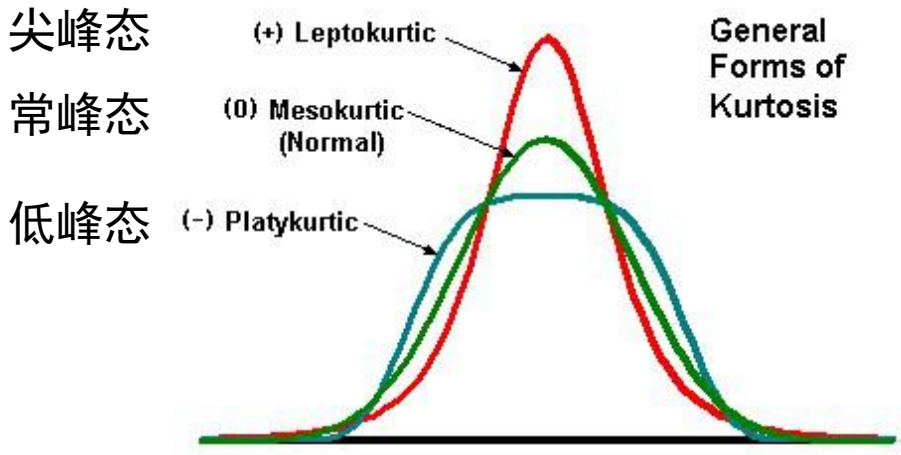
Central moments: $E\{[X - E(X)]^r\}$

1st ordinary moment: mean. 2nd central moment: variance.

3rd central moment: **Skewness**, asymmetry of a pdf.



4th central moment: **Kurtosis**, spikiness of a pdf.



The ***r*th moment** of a random variable, if exists, is $E(X^r)$.

Central moments: $E\{[X - E(X)]^r\}$

The mgf

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

$$M'(t) = \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f(x)\, dx = \int_{-\infty}^{\infty} x e^{tx} f(x)\, dx$$

$$M'(0) = \int_{-\infty}^{\infty} x f(x)\, dx = E(X)$$

Differentiating r times,

$$M^{(r)}(0) = E(X^r)$$

PROPERTY **B**

If the moment-generating function exists in an open interval containing zero, then $M^{(r)}(0) = E(X^r)$. ■

To find the moments we must sum a series or carry out an integration. But if mgf is found, the difficult process of integration or summation can be replaced by the mechanical process of differentiation!

Example:
Poisson distribution

$$M(t) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda}$$

$$= \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} e^{-\lambda}$$

$$= e^{-\lambda} e^{\lambda e^t}$$

$$= e^{\lambda(e^t - 1)}$$

Differentiating,

$$M'(t) = \lambda e^t e^{\lambda(e^t - 1)}$$

$$M''(t) = \lambda e^t e^{\lambda(e^t - 1)} + \lambda^2 e^{2t} e^{\lambda(e^t - 1)}$$

Evaluating them at t=0,

$$E(X) = \lambda$$

$$E(X^2) = \lambda^2 + \lambda$$

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \lambda$$

Mean and variance of a Poisson distribution are both λ.

Example:
Gamma distribution

$$M(t) = \int_0^\infty e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{x(t-\lambda)}\, dx$$

Gamma density with α, $\lambda - t$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{\Gamma(\alpha)}{(\lambda - t)^\alpha} \right) = \left(\frac{\lambda}{\lambda - t} \right)^\alpha$$

Differentiating,

$$M'(0) = E(X) = \frac{\alpha}{\lambda}$$

$$M''(0) = E(X^2) = \frac{\alpha(\alpha + 1)}{\lambda^2}$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

$$= \frac{\alpha(\alpha + 1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2}$$

$$= \frac{\alpha}{\lambda^2}$$

Example: **Normal distribution**

Standard normal distribution first:

$$M(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2}\, dx$$

配方大法好！

$$\frac{x^2}{2} - tx = \frac{1}{2}(x^2 - 2tx + t^2) - \frac{t^2}{2} = \frac{1}{2}(x-t)^2 - \frac{t^2}{2}$$

$$M(t) = \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2}\, dx \qquad u = x - t$$

Gaussian integrates to 1,

$$\boxed{M(t) = e^{t^2/2}}$$

Example: **Normal distribution**

Standard normal distribution first:

$$M(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2}\, dx$$

配方大法好！

$$\frac{x^2}{2} - tx = \frac{1}{2}(x^2 - 2tx + t^2) - \frac{t^2}{2} = \frac{1}{2}(x-t)^2 - \frac{t^2}{2}$$

$$M(t) = \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2}\, dx \qquad u = x - t$$

Gaussian integrates to 1,

$$M(t) = e^{t^2/2}$$

PROPERTY **C**

If X has the mgf $M_X(t)$ and $Y = a + bX$, then Y has the mgf $M_Y(t) = e^{at} M_X(bt)$.

Proof $\quad M_Y(t) = E(e^{tY}) = E(e^{at+btX}) = E(e^{at} e^{btX}) = e^{at} E(e^{btX}) = e^{at} M_X(bt)$

For a general Gaussian, $\quad M_Y(t) = e^{\mu t} M_X(\sigma t) = e^{\mu t} e^{\sigma^2 t^2/2}$

PROPERTY **D**

If X and Y are independent random variables with mgf's M_X and M_Y and $Z = X + Y$, then $M_Z(t) = M_X(t)M_Y(t)$ on the common interval where both mgf's exist.

Proof: $$M_Z(t) = E(e^{tZ}) = E(e^{tX+tY}) = E(e^{tX}e^{tY})$$

Independence, $$M_Z(t) = E(e^{tX})E(e^{tY}) = M_X(t)M_Y(t)$$

推广：多个独立随机变量，连乘即可。
矩母函数最有用的性质之一，可处理一些复杂的棘手问题。

Example: The sum of two independent Poisson random variables with parameters λ and μ is…?

PROPERTY **D**

If X and Y are independent random variables with mgf's M_X and M_Y and $Z = X + Y$, then $M_Z(t) = M_X(t)M_Y(t)$ on the common interval where both mgf's exist.

Proof: $$M_Z(t) = E(e^{tZ}) = E(e^{tX+tY}) = E(e^{tX}e^{tY})$$

Independence, $$M_Z(t) = E(e^{tX})E(e^{tY}) = M_X(t)M_Y(t)$$

推广：多个独立随机变量，连乘即可。
矩母函数最有用的性质之一，可处理一些复杂的棘手问题。

Example: The sum of two independent Poisson random variables with parameters λ and μ is…?

$$e^{\lambda(e^t-1)}e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)}$$

矩母函数**唯一决定**分布！

Example: The sum of independent normal random variables is…?

If $X \sim N(\mu,\ \sigma^2)$ and, independent of X, $Y \sim N(v,\ \tau^2)$, then the mgf of $X + Y$ is

$$e^{\mu t} e^{t^2\sigma^2/2} e^{vt} e^{t^2\tau^2/2} = e^{(\mu+v)t} e^{t^2(\sigma^2+\tau^2)/2}$$

which is the mgf of a normal distribution with mean $\mu + v$ and variance $\sigma^2 + \tau^2$. The sum of independent normal random variables is thus normal. ■

Example: The sum of independent normal random variables is…?

If $X \sim N(\mu, \sigma^2)$ and, independent of X, $Y \sim N(v, \tau^2)$, then the mgf of $X + Y$ is

$$e^{\mu t} e^{t^2\sigma^2/2} e^{vt} e^{t^2\tau^2/2} = e^{(\mu+v)t} e^{t^2(\sigma^2+\tau^2)/2}$$

which is the mgf of a normal distribution with mean $\mu + v$ and variance $\sigma^2 + \tau^2$. The sum of independent normal random variables is thus normal. ■

Example:

If X follows a gamma distribution with parameters α_1 and λ and Y follows a gamma distribution with parameters α_2 and λ, the mgf of $X + Y$ is

$$\left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1} \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_2} = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1+\alpha_2}$$

Note: Gamma density reduces to exponential when α_i=1, then $\alpha_1+\alpha_2+\ldots+\alpha_n=n$.

$$g(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t} \quad \rightarrow \quad \lambda e^{-\lambda t}$$

Sum of *n* independent exp. random variables with *λ* follows gamma with *n*, *λ*.

Poisson process: The time between n consecutive events in time follows gamma.

(e.g. Length of time to serve n customers in a queue)

***Note*:** *These cases are atypical, the sum may not follow the same distribution.*

Example: Random sums.
X_i are independent & with the same mgf M_X, N has M_N and indep. of X_i .

$$S = \sum_{i=1}^{N} X_i \qquad M_S(t) = E(e^{tS}) = E[E(e^{tS}|N)]$$

Given $N=n$, $$M_S(t) = [M_X(t)]^n$$

Thus,
$$\begin{aligned} M_S(t) &= E[M_X(t)^N] \\ &= E(e^{N \log M_X(t)}) \\ &= M_N[\log M_X(t)] \end{aligned}$$

The major limitation of the mgf is that it may not exist. The **characteristic function** of a random variable X is defined to be

$$\phi(t) = E(e^{itX})$$

Approximate methods

In many applications, only the 1st and 2nd moments of a random variable, instead of the entire probability distribution, are known *approximately.*

Why? Repeated independent observations of a random variable allow reliable estimates to be made of its *mean* and *variance*.

In many applications, only the 1st and 2nd moments of a random variable, instead of the entire probability distribution, are known *approximately.*

Why? Repeated independent observations of a random variable allow reliable estimates to be made of its *mean* and *variance*.

In reality, suppose we can measure X and determine its mean and variance, $Y=g(X)$, where g is a fixed function. How to determine $E(Y)$ and, at least approximately, Var (Y), in order to *assess the accuracy of the indirect measurement process*?

In many applications, only the 1st and 2nd moments of a random variable, instead of the entire probability distribution, are known *approximately.*

Why? Repeated independent observations of a random variable allow reliable estimates to be made of its *mean* and *variance*.

In reality, suppose we can measure X and determine its mean and variance, $Y=g(X)$, where g is a fixed function. How to determine $E(Y)$ and, at least approximately, Var (Y), in order to *assess the accuracy of the indirect measurement process*?

$E(Y)$ and Var(Y) cannot be calculated easily, unless g is linear. However, if g is nearly linear in a range in which X has high probability, it can be approximated by a linear function and approximate moments of Y can be found.

When confronted with a nonlinear problem hard to solve, we linearize.

In probability and statistics, this method is called **propagation of error**, or the **δ method.**

Taylor series expansion of g about μ_X, to the first order,

$$Y = g(X) \approx g(\mu_X) + (X - \mu_X)g'(\mu_X)$$

Approximately linear!

Recall: if $U=a+bV$, then $E(U)=a+bE(V)$ and $\mathrm{Var}(U)=b^2\mathrm{Var}(V)$,

$$\mu_Y \approx g(\mu_X) \qquad E(Y) \neq g(E(X))$$

$$\boxed{\sigma_Y^2 \approx \sigma_X^2[g'(\mu_X)]^2}$$

Expanding to 2nd order to improve it,

$$Y = g(X) \approx g(\mu_X) + (X - \mu_X)g'(\mu_X) + \tfrac{1}{2}(X - \mu_X)^2 g''(\mu_X)$$

Taking expectation of RHS, note $E(X - \mu_X) = 0$,

$$\boxed{E(Y) \approx g(\mu_X) + \tfrac{1}{2}\sigma_X^2 g''(\mu_X)}$$

How good such approximations are depends on how nonlinear g is in a neighborhood of μ_X and on the size of σ_X.

Example: The relation of voltage, current, and resistance is $V=IR$. Suppose that the voltage is held constant at V_0 across a medium whose resistance fluctuates randomly, say, of random fluctuations at the molecular level. The current therefore also varies randomly.

Suppose that it can be determined experimentally to have mean μ_I and variance σ_I^2. We wish to find the mean and variance of R.

$$R = g(I) = \frac{V_0}{I} \qquad g'(\mu_I) = -\frac{V_0}{\mu_I^2} \qquad g''(\mu_I) = \frac{2V_0}{\mu_I^3}$$

$$\mu_R \approx \frac{V_0}{\mu_I} + \frac{V_0}{\mu_I^3}\sigma_I^2$$

$$\sigma_R^2 \approx \frac{V_0^2}{\mu_I^4}\sigma_I^2$$

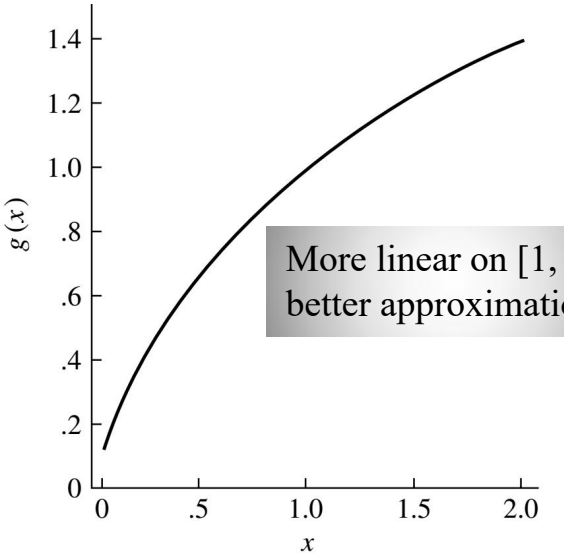
- σ_R depends on both mean and variance of I.
- For small I, change of I leads to large variations in $R=V_0/I$.
- For small I, 2nd order correction for μ_R is large.
- When $I \to 0$, function is highly non-linear, not a good approximation.

Example: Test the accuracy of the approximations. $g(x) = \sqrt{x}$
Consider two cases: *X* uniform on [0, 1]; *X* uniform on [1, 2].

Exact result. Let $Y = \sqrt{X}$, for X uniform on [0, 1],

$$E(Y) = \int_0^1 \sqrt{x}\, dx = \tfrac{2}{3} \qquad E(Y^2) = \int_0^1 x\, dx = \tfrac{1}{2}$$

$$\text{Var}(Y) = \tfrac{1}{2} - \left(\tfrac{2}{3}\right)^2 = \tfrac{1}{18} \text{ and } \sigma_Y = .236.$$



More linear on [1, 2], better approximation?

Approx. X uniform on [0, 1], μ_X=1/2, Var (X)=1/12,

$$g'(x) = \tfrac{1}{2}x^{-1/2} \qquad g'(\mu_X) = \frac{\sqrt{2}}{2}$$

$$g''(x) = -\tfrac{1}{4}x^{-3/2} \qquad g''(\mu_X) = -\frac{\sqrt{2}}{2}$$

$$E(Y) \approx \sqrt{\frac{1}{2} - \frac{1}{2}\left(\frac{\sqrt{2}}{12 \times 2}\right)} = .678$$

$$\text{Var}(Y) \approx \tfrac{1}{2} \times \tfrac{1}{12} = .042$$

$$\sigma_Y \approx .204$$

0.667 (1.6%)

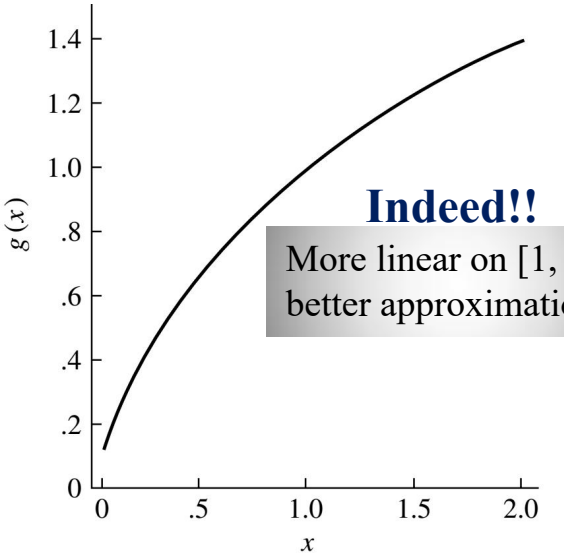
0.236 (13%)

$$E(Y) \approx g(\mu_X) + \tfrac{1}{2}\sigma_X^2 g''(\mu_X)$$

$$\sigma_Y^2 \approx \sigma_X^2 [g'(\mu_X)]^2$$

Example: Test the accuracy of the approximations. $g(x) = \sqrt{x}$
Consider two cases: X uniform on [0, 1]; X uniform on [1, 2].

Exact result. For X uniform on [1, 2], $Y = \sqrt{X}$
Mean=1.219, Var(Y)= .142, $\sigma_Y = .119$.



Indeed!!
More linear on [1, 2], better approximation?

Approx. X uniform on [1, 2], μ_X=3/2, Var (X)=1/12,

$$g'(\mu_X) = .408$$
$$g''(\mu_X) = -.136$$

$$E(Y) \approx g(\mu_X) + \tfrac{1}{2}\sigma_X^2 g''(\mu_X)$$
$$\sigma_Y^2 \approx \sigma_X^2 [g'(\mu_X)]^2$$

$$E(Y) \approx \sqrt{\frac{3}{2}} - \frac{1}{2}\left(\frac{.136}{12}\right) = 1.219$$
$$\mathrm{Var}(Y) \approx \frac{.408^2}{12} = .0138$$
$$\sigma_Y \approx .118$$

1.219 (0%)

0.119 (1%)

The case of 2-variables

$$Z = g(X, Y)$$

Taylor series expansion of g about (μ_X, μ_Y) – denoted as μ – to 1st order,

$$Z = g(X, Y) \approx g(\mu) + (X - \mu_X)\frac{\partial g(\mu)}{\partial x} + (Y - \mu_Y)\frac{\partial g(\mu)}{\partial y}$$

Approximately linear!

$$E(Z) \approx g(\mu)$$

$\sigma_Y^2 \approx \sigma_X^2[g'(\mu_X)]^2$

$$\mathrm{Var}(Z) \approx \sigma_X^2\left(\frac{\partial g(\mu)}{\partial x}\right)^2 + \sigma_Y^2\left(\frac{\partial g(\mu)}{\partial y}\right)^2 + 2\sigma_{XY}\left(\frac{\partial g(\mu)}{\partial x}\right)\left(\frac{\partial g(\mu)}{\partial y}\right)$$

Expanding to 2nd order for an improved estimate,

$$Z = g(X, Y) \approx g(\mu) + (X - \mu_X)\frac{\partial g(\mu)}{\partial x} + (Y - \mu_Y)\frac{\partial g(\mu)}{\partial y}$$
$$+ \frac{1}{2}(X - \mu_X)^2\frac{\partial^2 g(\mu)}{\partial x^2} + \frac{1}{2}(Y - \mu_Y)^2\frac{\partial^2 g(\mu)}{\partial y^2}$$
$$+ (X - \mu_X)(Y - \mu_Y)\frac{\partial^2 g(\mu)}{\partial x \partial y}$$

Taking expectations of RHS,

$E(Y) \approx g(\mu_X) + \frac{1}{2}\sigma_X^2 g''(\mu_X)$

$$E(Z) \approx g(\mu) + \frac{1}{2}\sigma_X^2\frac{\partial^2 g(\mu)}{\partial x^2} + \frac{1}{2}\sigma_Y^2\frac{\partial^2 g(\mu)}{\partial y^2} + \sigma_{XY}\frac{\partial^2 g(\mu)}{\partial x \partial y}$$

Example: *Expectation and variance of a ratio.*

A chemist measures the concentrations of two substances, both with some measurement error that is indicated by their standard deviations, and then report the relative concentrations in the form of a ratio.

What is the approximate standard deviation of the ratio, *Z=Y/X* ?

For $g(x, y) = y/x$,

$$\frac{\partial g}{\partial x} = \frac{-y}{x^2} \quad \frac{\partial g}{\partial y} = \frac{1}{x}$$

$$\frac{\partial^2 g}{\partial x^2} = \frac{2y}{x^3} \quad \frac{\partial^2 g}{\partial y^2} = 0 \quad \frac{\partial^2 g}{\partial x \partial y} = \frac{-1}{x^2}$$

$$E(Z) \approx g(\mu) + \frac{1}{2}\sigma_X^2 \frac{\partial^2 g(\mu)}{\partial x^2} + \frac{1}{2}\sigma_Y^2 \frac{\partial^2 g(\mu)}{\partial y^2} + \sigma_{XY} \frac{\partial^2 g(\mu)}{\partial x \partial y}$$

$$E(Z) \approx \frac{\mu_Y}{\mu_X} + \sigma_X^2 \frac{\mu_Y}{\mu_X^3} - \frac{\sigma_{XY}}{\mu_X^2} = \frac{\mu_Y}{\mu_X} + \frac{1}{\mu_X^2}\left(\sigma_X^2 \frac{\mu_Y}{\mu_X} - \rho\sigma_X\sigma_Y\right)$$

Small if σ_X, σ_Y are small, measured accurately; large if μ_X is small.

$$\mathrm{Var}(Z) \approx \sigma_X^2 \left(\frac{\partial g(\mu)}{\partial x}\right)^2 + \sigma_Y^2 \left(\frac{\partial g(\mu)}{\partial y}\right)^2 + 2\sigma_{XY} \left(\frac{\partial g(\mu)}{\partial x}\right)\left(\frac{\partial g(\mu)}{\partial y}\right)$$

$$\mathrm{Var}(Z) \approx \sigma_X^2 \frac{\mu_Y^2}{\mu_X^4} + \frac{\sigma_Y^2}{\mu_X^2} - 2\sigma_{XY} \frac{\mu_Y}{\mu_X^3} = \frac{1}{\mu_X^2}\left(\sigma_X^2 \frac{\mu_Y^2}{\mu_X^2} + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y \frac{\mu_Y}{\mu_X}\right)$$

Quite variable if μ_X is small. Var(Z) decreases if ρ and μ_Y/μ_X are of the same sign.

Laplace's law of succession

Suppose that the sun has risen n times in succession; what is the probability that it will rise once more?

Laplace used an "urn model" to study successive sunrise as a random process. A sunrise is assimilated to the drawing of a black ball from an urn of unknown composition. The various possible compositions are assimilated to so many different urns containing various proportions of black balls. Finally, the choice of the true value of the proportion is assimilated to the picking of a random number in [0, 1]. Clearly, these are weighty assumptions calling forth serious objections at several levels.

- Is sunrise a random phenomenon or is it deterministic?
- Assuming that it can be treated as random, is the preceding simple urn model adequate to its description?
- Assuming that the model is appropriate in principle, why should the a priori distribution of the true probability be uniformly distributed, and if not how could we otherwise assess it?

Problem set #2

Suppose that in a numerical simulation, you need to generate some fake noise that follows a **standard normal distribution**. Since the cdf has no closed form, let us do it in two ways:

1. *Rejection method*. (cf. Lec 3, pp. 28). Although a Gaussian is defined on $(-\infty, \infty)$, you can approximate it by a large enough interval (e.g. $[-3, 3]$, $[-5, 5]$), and choose $m(x)$ to be uniformly distributed. Plot a histogram using your output data.

2. *Polar method*. In Lec 4 we find that if X, Y are Gaussian, Θ is uniform on $[0, 2\pi]$, R has a Rayleigh density. Let $T=R^2$, by calculating the cdf (Lec 2, pp. 32), we obtain

$$f_T(t) = \frac{1}{2}e^{-t/2}, \qquad t \geq 0 \qquad\qquad f_{T\Theta}(t, \theta) = \frac{1}{2\pi}\left(\frac{1}{2}\right)e^{-t/2}$$

where the joint density is due to the fact that T, Θ are also independent. R^2 is exponential!

Therefore, we can the do following: First, you generate independent random variables U_1 and U_2 , both uniformly distributed on $[0, 1]$. Then $-2 \log U_1$ is exponentially distributed with parameter ½, and $2\pi\, U_2$ is uniform on $[0, 2\pi]$. Therefore, the following X, Y are independent standard normal random variables. Again, plot a histogram demonstrating that your output indeed follows $N(0, 1)$.

$$X = \sqrt{-2\log U_1}\cos(2\pi U_2) \qquad\qquad Y = \sqrt{-2\log U_1}\sin(2\pi U_2)$$