Functions of random variables

- **X** is a random variable
- Let **Y** = g(**X**) where g(•) is some specified function
- What is the probabilistic behavior of **Y**?
- Some answers are readily obtained
- LOTUS tells us how to find E[**Y**] = E[g(**X**)] from the pmf/pdf of **X**
- LOTUS also gives E[$\mathbf{Y}^2$] = E[g^2(**X**)] from which we get var(**Y**) = E[g^2(**X**)] – E^2[g(**X**)]

What kind of random variable is Y?

- If **X** is a discrete random variable taking on values u_1, u_2, … , u_n, …, then **Y** must also a discrete random variable
- **Y** takes on values g(u_1), g(u_2), … g(u_n), …
- Note that some values may be the same, i.e., g(u_i) = g(u_j) for some choices of i and j
- The pmf of **Y** is readily obtained from this
- $p_{\mathbf{Y}}(v_j)$ = P{**Y** = v_j} = ∑$p_{\mathbf{X}}(u_i)$ where the sum is over all i such that g(u_i) = v_j

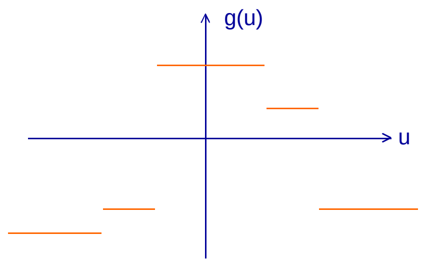
But what if X is a continuous RV?

- If **X** is a continuous random variable, then depending on the function g(•), **Y** may be a
  - discrete random variable
  - continuous random variable
  - mixed random variable
  - singular random variable
- Technicality: g(•) must be measurable, but we only deal with measurable functions

When will Y be a discrete RV?

- Let **X** be a continuous random variable
- Suppose that the real numbers can be partitioned into a (countable) set of intervals I_1, I_2, … I_n, … such that g(u) has constant value v_j for all u I_j
- Then, **Y** takes values v_1, v_2, … v_n, … with probabilities P{**X** I_1}, P{**X** I_2}, … , P{**X** I_n}, … respectively

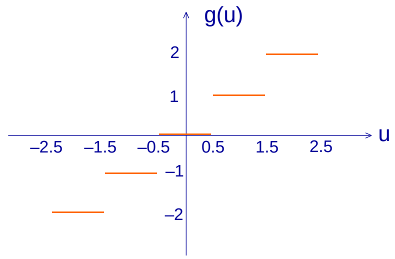
An example of such a function



A quantizer function

- Example: g(•) is a staircase function or quantizer specified as follows:
- For each value of u, let n be the unique integer such that n–0.5 < u ≤ n+0.5. Then, g(u) = n, that is, g(u) has constant (integer) value n for all u (n–0.5, n+0.5]
- A graph of what the function g(•) looks like is shown next

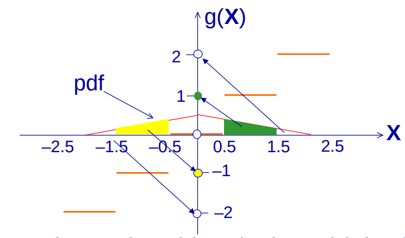
The quantizer function g(u)



Continuous to discrete

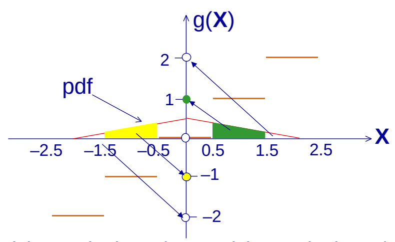
- Example: g(u) has constant (integer) value n for all u (n–0.5, n+0.5]
- Thus, if the value of **X** on any trial of the experiment is in the interval (n–0.5, n+0.5], then the value of Y on that trial is n
- $p_Y(n)$ = P{**Y** = n} = P{**X** (n–0.5, n+0.5]}
  = area under pdf curve $f_X(u)$
  from u = n–0.5 to u = n+0.5
  = $F_X(n+0.5) - F_X(n-0.5)$

How to think about it …



Mass is transferred from horizontal (**X**) axis to vertical (**Y**) axis

What's the final answer?



$p_Y(2) = p_Y(-2) = 1/32$; $p_Y(1) = p_Y(-1) = 1/4$;
$p_Y(0) = 7/16$

Continuous to discrete (continued)

- Example: g(u) has constant (integer) value n for all u (n–0.5, n+0.5]
- $p_Y(n) = F_X(n+0.5) - F_X(n-0.5)$
- Let **X** be a $N(\mu, \sigma^2)$ RV
- Then, $p_Y(n)$= P{**X** (n–0.5, n+0.5]}
  = $\Phi((n+0.5-\mu)/\sigma) - \Phi((n-0.5-\mu)/\sigma)$
  which can be found from tables …
  or by using your calculator

Round up all the usual suspects!

- Example: h(•) has constant (integer) value n on the interval (n–1, n]. Thus, **Z** = h(**X**) = éXù rounds up the value of **X**
- $p_Z(n)$ = P{**X** (n–1, n]}
- If **X** is an exponential RV with parameter λ then, for n ≥ 1, $p_Z(n)$ = P{**X** (n–1, n]}
  = $\exp(-\lambda(n-1)) - \exp(-\lambda n)$
  = $\exp(-\lambda(n-1))\bullet[1 - \exp(-\lambda n)]$
  = $[\exp(-\lambda)]^{n-1}\bullet[1 - \exp(-\lambda)]$ = $q^{n-1}\bullet p$

Exponential RV → Geometric RV

- If **X** is an exponential RV with parameter λ, **Z** = é**X**ù is a geometric random variable with parameter p = 1 – exp(–λ)
- $p_Z(n) = [1 - \exp(-\lambda)] \cdot [\exp(-\lambda)]^{n-1}$ for $n \geq 1$
- Geometric RVs and exponential RVs are memoryless: $P\{\mathbf{X} > a+b \mid \mathbf{X} > a\} = P\{\mathbf{X} > b\}$
- So, here is another analogy between them based on rounding up the value
- What about waiting times?

Rounding up waiting times

- Clock signal at t = 0, 1, 2, …
- Poisson process with arrival rate λ/clock
- Arrivals (if any) during the time interval (n–1, n] cannot be noticed till the next clock signal at t = n
- Waiting time for the next non-empty clock interval is a geometric RV with parameter p = 1 – exp(–λ). P{no arrival} = exp(–λ)
- Note: P{at least one arrival} = 1 – exp(–λ)

Alas, the easy stuff is all done …

- discrete → discrete transformations and continuous → discrete transformations of random variables are easy to handle
- continuous → continuous transformations

  and

  continuous → mixed transformations

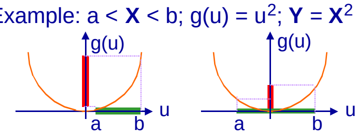
  of random variables require more careful work, but, as usual, DAD can help a lot

Continuous RV → Continuous RV

- Let **X** denote a continuous RV and g(u) a continuous function of u such that g´(u) is not zero on any interval
  - Note: Having g´(u) = 0 for some values of u is acceptable, but g´(u) = 0 for all u such that s < u < t is not allowed
  - This restriction avoids "flat spots" in g(•) that give rise to mixed or discrete RVs
- Then, **Y** = g(**X**) is a continuous RV

Y = g(X) is a continuous RV

- In order to determine the pdf of **Y**, it is very useful to first draw a sketch of g(u)
- The sketch makes it easy to determine the range of values that **Y** can take on
- Example: $a < \mathbf{X} < b$; $g(u) = u^2$; $\mathbf{Y} = \mathbf{X}^2$



- $0 < a$ $a^2 < \mathbf{Y} < b^2$
- $a<0<b$ $0 \leq \mathbf{Y} < \max\{a^2,b^2\}$

Ask DAD for help! Draw a diagram!

- Sketch the function g(u) and mark the range of values of **X** on the horizontal axis
- This readily allows you to figure out the range of values (say, from s to t) that **Y** will take on
- Thus, you can immediately write down
  - $f_Y(v)$ = ??, for $s < v < t$
  - $f_Y(v) = 0$ elsewhere
- I got 3 of 4 parts right. Worth at least 75% partial credit?

Continuing the good work…

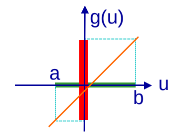
- There are two ways to proceed from here
- Method I: Figure out the CDF $F_Y(v)$ and then differentiate the CDF to get $f_Y(v)$
- Advantages:
  - already know that $F_Y(v) = 0$ for $v < s$; $F_Y(v) = 1$ for $v > t$
  - method always works
  - if you do it right!
- Method II: Use a mystical magical formula

Trying Plan B instead …

- Method II: Use the mystical magical formula in Theorem 7.1 (p. 239) of Ross
- Advantages:
  - when it works, it is easy to get the right answer
  - if you use it right, of course!
- Disadvantage:
  - method doesn't always work as simply as the theorem seems to imply

What's Plan A (a.k.a Method I) like?

- Example: **X** takes on values in the range (a, b), and has pdf $f_X(u)$. **Y** = c•**X** + d, where c > 0. What is $f_Y(v)$?



- **Y** takes on values between s = c•a + d and t = c•b + d

Method I (continued)

- **Y** takes on values between s = c•a + d and t = c•b + d
- The CDF $F_Y(v) = 0$ for $v < s$ and 1 for $v > t$
- For any number v such that s < v < t,

  $F_Y(v) = P\{\mathbf{Y} \le v\} = P\{c\cdot\mathbf{X} + d \le v\}$

  $= P\{\mathbf{X} \le (v–d)/c\} = F_X((v–d)/c)$
- Hence, for s<v<t, $f_Y(v)$ = derivative of $F_Y(v)$

= derivative of $F_X((v–d)/c) = (1/c)\cdot f_X((v–d)/c)$

Method I (continued)

- For s<v<t, $f_Y(v)$ = derivative of $F_Y(v)$

= derivative of $F_X((v–d)/c) = (1/c)\cdot f_X((v–d)/c)$

- Since $F_Y(v)$ is constant for v < s or v > t,

  $f_Y(v)$ = derivative of $F_Y(v)$

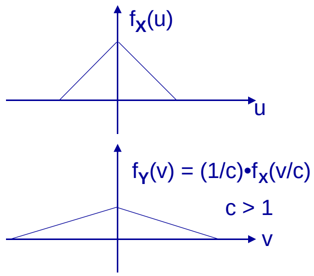
  = 0 in these regions
- $f_Y(v) = (1/c)\cdot f_X((v–d)/c)$ holds for all real numbers v since if v > t = c•b +d, then (v–d)/c > b $f_X((v–d)/c) = 0$, etc.

Linear transformations

- **Y** = c•**X** + d is a linear function of **X**
- Assume that d = 0. Then,

  $f_Y(v) = (1/c)\cdot f_X((v–d)/c) = (1/c)\cdot f_X(v/c)$
- The function $f_X(v/c)$ is just $f_X(v)$ stretched out by a factor of c (c > 1) or squeezed in by a factor of c (c < 1)
- If horizontal scale expands by a factor c, the vertical scale must compress by the same factor to keep the area the same

A graphical interpretation



Linear transformations (continued)

- **Y** = c•**X** is a linear function of **X**
- $f_Y(v) = (1/c)•f_X(v/c)$
- The general shape of $f_Y(v)$ is the same as the general shape of $f_X(u)$, but the horizontal and vertical scales are scaled inversely
- The constant term d just translates the pdf rightward by d
- Exercise: repeat for the case c < 0

Linear transformations (continued)

- **Y** = c•**X** is a linear function of **X**
- $f_Y(v) = (1/c)•f_X(v/c)$
- The general shape of $f_Y(v)$ is the same as that of $f_X(u)$, but the horizontal and vertical axes are scaled inversely
- Example: Gamma RV **Y** with parameters (t, λ) is just **X**/λ where **X** is a gamma RV with parameters (t, 1)
- λ is the scale parameter

Gaussian random variables

- When **X** is a Gaussian RV, **Y** = c•**X** + d is also a Gaussian RV (with different mean and variance)
- For c > 0, $F_Y(v) = F_X((v–d)/c)$
- If **X** is $N(\mu, \sigma^2)$, $F_X(u) = \Phi((u – \mu)/\sigma)$
- $F_Y(v) = F_X((v–d)/c) = \Phi(((v–d)/c – \mu)/\sigma)$
  $= \Phi(v– c•\mu–d)/c•\sigma) = \Phi(v– (c•\mu+d))/c•\sigma)$
- **Y** is $N(c•\mu+d, c^2•\sigma^2)$

Some remarks on Gaussianity

- Given any arbitrary RV **X** with E[**X**] = μ and var(**X**) = σ^2, E[**Y**] = E[c•**X** + d] = c•μ+d and var(**Y**) = $c^2•\sigma^2$
- This holds for Gaussian RVs **X** also
- But, more important, **Y** also happens to be Gaussian with this mean and variance
- **Y** ~ $N(c•\mu+d, c^2•\sigma^2)$
- We can write down the pdf of **Y** directly after calculating its mean and variance

Even more remarks on Gaussianity

- That linear functions of Gaussian RVs are Gaussian RVs applies very generally
  - linear functions of multiple Gaussian RVs are Gaussian RVs
  - The output of a linear system is a Gaussian random process if its input is a Gaussian random process
- GIGO = Gaussian In Gaussian Out
- var(output)
  = var(input)•(energy in impulse response)

Other applications of Method A

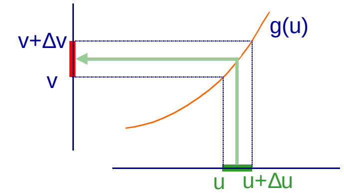
- Ross shows that if $\mathbf{Y} = \mathbf{X}^2$, then for $v \geq 0$
  $f_Y(v) = [f_X(\sqrt{v}) + f_X(-\sqrt{v})]/(2\bullet\sqrt{v})$
- If $\mathbf{X} \sim N(0, 1)$, then
  $f_Y(v) = (2\pi v)^{-1/2}\bullet\exp(-v/2)$
  which is a gamma pdf with parameters (1/2, 1/2), i.e., a chi-square pdf with one degree of freedom
- If $\mathbf{X} \sim N(0, \sigma^2) = \sigma\bullet N(0, 1)$, the gamma pdf parameters are $(1/2, 1/(2\bullet\sigma^2))$

How Method B is applied

- The mystical magical formula in Ross's Theorem 7.1 (p. 239) works as follows:
  - $v = g(u) = c\bullet u + d$, so $g^{-1}(v) = (v-d)/c$
  - derivative of $g^{-1}(v)$ is $1/c$
  - $f_Y(v) = f_X(g^{-1}(v))\bullet|\text{derivative of } g^{-1}(v)|$
    $= |1/c|\bullet f_X((v-d)/c)$
- The mystical magical formula gives the answer for $c > 0$ as well as for $c < 0$

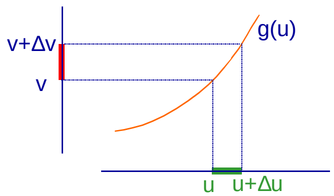
Why does Method B work, anyway?

- The mass in the interval $(u, u+\Delta u)$ on the horizontal (**X**) axis is transferred to the interval $(v, v+\Delta v)$ on the vertical (**Y**) axis



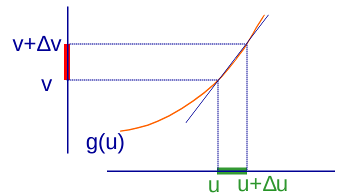
How much mass is transferred?

- Mass in the interval $(u, u+\Delta u) \approx f_X(u)\bullet\Delta u$
  Mass in the interval $(v, v+\Delta v) \approx f_Y(v)\bullet\Delta v$
- $f_Y(v)\bullet\Delta v = f_X(u)\bullet\Delta u$. Note that $g(u) = v$



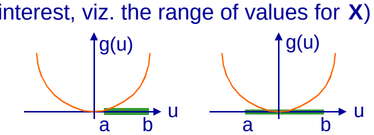
Putting it all together

- $f_Y(v)\bullet\Delta v = f_X(u)\bullet\Delta u$
- $g(u) = v$, so $u = g^{-1}(v)$
- $\Delta v/\Delta u = g'(u)$
- $\Delta u/\Delta v = [g^{-1}(v)]'$
- $f_Y(v) = f_X(g^{-1}(v))\bullet|[g^{-1}(v)]'|$



When does Method B fail?

- Method B requires that $g(u)$ be monotone increasing or monotone decreasing for all u, (or at least for all u in the range of interest, viz. the range of values for **X**)



- Method B works for the left-hand case but not for the right-hand case

Can Method B be generalized?

- Yes, there is a more general version of Method B that does not require g(u) to be a monotone function
- But the details are monotonous …
- And the method requires so much extra work in its applications that many times, one is, in effect, back to Method A
- Do not rely on Method B; it is strictly a fair-weather friend!

A special application

- The mystical magical formula says that $f_Y(v) = f_X(g^{-1}(v))\bullet[g^{-1}(v)]' = f_X(g^{-1}(v))/g'(u)$
- Now suppose that we choose $g(u) = F_X(u)$, i.e. g(u) is the CDF of **X**
- Why not? The CDF of **X** is a nice well-behaved monotone increasing function!
- **Y** = F_X(**X**) takes on values in [0, 1]
- So, we only need to find $f_Y(v)$ for $0 < v < 1$

Pulling the rabbit out of the hat

- Remember that g(u) = v and u = $g^{-1}(v)$
- For any v, $0 < v < 1$
  $f_Y(v) = f_X(g^{-1}(v))\bullet[g^{-1}(v)]' = f_X(g^{-1}(v))/g'(u)$
  $= f_X(u)/g'(u)$
- If $g(u) = F_X(u)$, then $g'(u) = f_X(u)$, right?
- So, for any v between 0 and 1, $f_Y(v) = 1$!!
- **Y** = F_X(**X**) is uniformly distributed on (0, 1)
- Applying the CDF as a function to a RV **X** results in a uniform RV on (0, 1)

Of great use in simulation

- Applying the CDF as a function to a RV **X** results in a uniform RV **Y** on (0, 1)
- Suppose that we wish to create an RV **X** with a specified CDF F(•)
- Given RV **Y** uniformly distributed on (0, 1), we can apply the inverse function $F^{-1}(\bullet)$ to **Y** to get an RV **X** with CDF F(•)
- This technique is often used to simulate an RV **X** using the output of `rand()`

An example of simulation

- Let **Y** be uniformly distributed on (0, 1)
- We want **X** to be an exponential RV with parameter 1: $F(u) = 1 - \exp(-u)$ for $u \geq 0$
- For $0 < v < 1$, $F^{-1}(v) = -\ln(1-v)$
- Thus, – ln(1– **Y**) is an exponential RV with parameter 1. So is – ln(**Y**). Why?
- – ln(`rand()`) simulates an exponential RV RV **X** with parameter 1
- What does – ln (`rand()`)/λ simulate?

Summary

- We have discussed the methods used to compute the pmf or pdf of **Y** = g(**X**)
- If **X** is discrete, so is **Y**
- When **X** is continuous, **Y** can be discrete
- When **Y** is continuous, $F_Y(v)$ is found, and then differentiated to get $f_Y(v)$
- A magical formula sometimes gives easy answers, but requires care in its use
- **X** = F^{-1}(`rand()`) simulates **X**