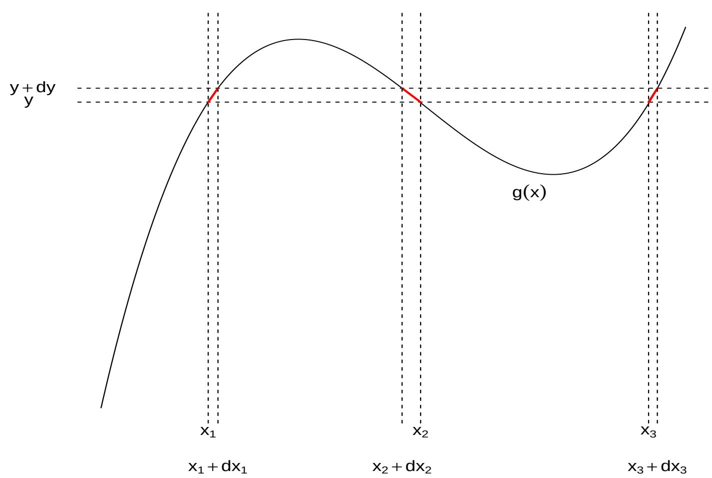
Class Notes 3-13-2019

- **General Formula for Density of $Y = g(X)$ when g is not 1-1**

$$f_Y(y) = \sum_{\substack{x:g(x)=y \\ g'(x)\neq 0}} f_X(x)\frac{1}{|g'(x)|}$$

Here we do not replace x in $f_X(x)$ and $g'(x)$ by $x = g^{-1}(y)$ because g does not necessarily have an inverse. There may be several values x that map into the same y. That is the reason for the notation used.

The above formula for $f_Y(y)$ can easily be derived via the infinitesimal interpretation of $f_Y(y)dy$ as approximation to $P(Y \in [y, y+dy])$, i.e., $P(Y \in [y, y+dy]) \approx f_Y(y)dy$



In the plot the x_i abscissa values correspond to the intersect locations of the non-monotone function with the horizontal y line, i.e., $g(x_i) = y$. The intercepts $x_i + dx_i$ correspond to the corresponding locations of the intercepts with the $y + dy$ line, i.e., $g(x_i + dx_i) = y + dy$. The thick red secants linearly connect (x_i, y) with $(x_i + dx_i, y + dy)$ with slope approximately $g'(x_i) \approx \frac{dy}{dx_i}$. We then have in this particular case (with 3 intersects, i.e., 3 x's with $g(x) = y$),

$$P(Y \in [y, y+dy]) = P(X \in [x_1, x_1 + dx_1]) + P(X \in [x_2 + dx_2, x_2]) + P(X \in [x_3, x_3 + dx_3])$$

where the reversal of endpoints in the interval $[x_2 + dx_2, x_2]$ results from the fact the $g(x)$ is decreasing over that interval, i.e., $dx_2 < 0$. Using the infinitesimal interpretation of such probabilities we have

$$P(X \in [x_i, x_i + dx_i] \approx f_X(x_i)dx_i, \quad \text{for } i = 1, 3 \text{ and} \quad P(X \in [x_2 + dx_2, x_2]) \approx -f_X(x_2)dx_2,$$

Hence

$$f_Y(y)dy \approx P(Y \in [y, y+dy]) \approx f_X(x_1)dx_1 - f_X(x_2)dx_2 + f_X(x_3)dx_3$$

or, after dividing both sides by dy

$$f_Y(y) \approx f_X(x_1)\frac{1}{\frac{dy}{dx_1}} - f_X(x_1)\frac{1}{\frac{dy}{dx_2}} + f_X(x_3)\frac{1}{\frac{dy}{dx_3}} \longrightarrow f_X(x_1)\frac{1}{|g'(x_1)|} + f_X(x_2)\frac{1}{|g'(x_2)|} + f_X(x_3)\frac{1}{|g'(x_3)|}$$

as all the infinitesimals become arbitrarily small, resulting in the above more general expression in this case.

- **Revisiting $Y = g(Z) = Z^2$ with $Z \sim \mathcal{N}(0,1)$:**
  Since $Y = g(Z)$ we use the symbol z in place of x in our formula above. We have $g'(z) = 2z$ and $z^2 = g(z) = y$ has two solutions for $y > 0$, namely $z_1 = \sqrt{y}$ and $z_2 = -\sqrt{y}$. Thus for $y > 0$ our formula for $f_Y(y)$ becomes

$$f_Y(y) = f_Z(z_1)\frac{1}{|2z_1|} + f_Z(z_2)\frac{1}{|2z_2|} = \varphi(\sqrt{y})\frac{1}{2\sqrt{y}} + \varphi(-\sqrt{y})\frac{1}{2\sqrt{y}} = \varphi(\sqrt{y})\frac{1}{\sqrt{y}}$$

  exactly the same expression we got via the cdf approach and differentiation.

- **Some Special Transforms Involving the CDF $F(x)$ of X**
  Let us take the standard definition of the inverse cdf $F^{-1}(u)$, namely

$$F^{-1}(u) = \inf\{x : u \le F(x)\} \qquad \text{and note that} \qquad F^{-1}(u) \le x \iff u \le F(x)$$

  so that for $U \sim \text{Unif}(0,1)$ we have $P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x)$ for all $x \in \mathbb{R}$.
  This gives us a recipe for generating random variables $X \sim F$ once we have a way of generating $U \sim \text{Unif}(0,1)$ and if we know how to compute the inverse $F^{-1}(u)$.
  We note that $F^{-1}(p)$ is not the same definition as given previously for the p-quantile of a distribution. When $F(x) = p$ happens to occur over an interval $[a,b)$ then any $x \in [a,b)$ would qualify as a p-quantile. While $F^{-1}(p)$ is also a p-quantile, it is a unique choice.

  Based on the above result about the cdf of $F^{-1}(U)$ one might suspect that $U = F(X)$ is distributed like $\text{Unif}(0,1)$. That is not the case in general and can easily be seen for discrete random variables. The simplest case is that of a constant random variable, e.g., $P(X = 0) = 1$. Then $F(X)$ can take on only the value $F(X) = F(0) = P(X \le 0) = 1$ while $P(X < 0) = 0$ so that $P(F(X) = 1) = 1$, i.e., $F(X) \not\sim \text{Unif}(0,1)$.
  However, $F(X) \sim \text{Unif}(0,1)$ when the cdf $F(x)$ is continuous.
  This is seen as follows: If for a given $u \in (0,1)$ we have that $u = F(x)$ at a unique point $x = F^{-1}(u)$, i.e., F crosses the level u in strictly monotone increasing fashion.
  Then $P(F(X) \le u) = P(X \le F^{-1}(u)) = F(F^{-1}(u)) = u$.
  If we have $F(x) = u$ for all $x \in [a,b]$ and $F(x) < u$ for $x < a$ and $F(x) > u$ for $x > b$ then

$$P(F(X) \le u) = P(X \le x) = u \quad \text{for all } x \in [a,b]$$

  Where does continuity of F come in?
  We need a value x where $F(x) = u$, i.e., $F(x)$ cannot jump over the level u.

  This latter result that $F(X) \sim \text{Unif}(0,1)$ when F is continuous is extremely important in the study of large sample theory (asymptotics) because is allows us to reduce many results concerning iid $X_1, \ldots, X_n \sim F$ to the study of iid $U_1, \ldots, U_n \sim \text{Unif}(0,1)$, i.e., we don't have to prove theorems over and over for different F.

- **The Sum of Independent Poisson Random Variables:**
  Theorem 1: Let X_1 and X_2 be independent Poisson random variables with respective parameters $\lambda_1 > 0$ and $\lambda_2 > 0$. Then $S = X_1 + X_2$ is a Poisson random variable with parameter $\lambda_1 + \lambda_2$.

Proof:

$$\begin{aligned}
P\left(X_1+X_2=z\right) &= \sum_{i=0}^{\infty} P\left(X_1+X_2=z, X_2=i\right)=\sum_{i=0}^{\infty} P\left(X_1+i=z, X_2=i\right) \\
&= \sum_{i=0}^{z} P\left(X_1=z-i, X_2=i\right)=\sum_{i=0}^{z} \frac{e^{-\lambda_1} \lambda_1^{z-i}}{(z-i)!} \frac{e^{-\lambda_2} \lambda_2^{i}}{i!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^z}{z!} \sum_{i=0}^{z} \frac{z!}{i!(z-i)!}\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{i}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{z-i} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^z}{z!}
\end{aligned}$$

Corollary: If $X_1, \ldots, X_n$ are independent Poisson random variables with respective parameters $\lambda_1, \ldots, \lambda_n$ then $S = X_1+\ldots+X_n$ is a Poisson random variable with parameter $\lambda_1+\ldots+\lambda_n$.

Proof: By induction over n.

$$X_1+\ldots+X_{n-1} \sim \text{Pois}(\lambda_1+\ldots+\lambda_{n-1}) \quad \text{and} \quad X_n \sim \text{Pois}(\lambda_n)$$
$$\Longrightarrow \quad X_1+\ldots+X_n \sim \text{Pois}(\lambda_1+\ldots+\lambda_n)$$

Chapter 9: Tail Bounds, Section 9.1-9.2

- **Markov's Inequality:**

$$P(|X| \geq c) \leq \frac{E(|X|)}{c} \qquad \text{for any } c>0$$

Proof: $|X| = |X|I_{\{|X|\geq c\}} + |X|I_{\{|X|<c\}} \geq |X|I_{\{|X|\geq c\}} \geq cI_{\{|X|\geq c\}}$. Thus

$$E(|X|) \geq cE\left(I_{\{|X|\geq c\}}\right) = cP(|X| \geq c)$$

- **Chebychev's Inequality:**
Let X be a random variable with finite mean $\mu = E(X)$ and finite variance $\sigma^2 = \text{var}(X)$. Then for any $c > 0$ we have

$$P\left(|X-\mu| \geq c\right) \geq \frac{\sigma^2}{c^2}$$

Proof: Using Markov's inequality on $||X-\mu|^2| = |X-\mu|^2$ we have

$$P\left(|X-\mu| \geq c\right) = P\left(|X-\mu|^2 \geq c^2\right) \leq \frac{E(|X-\mu|^2)}{c^2} = \frac{\sigma^2}{c^2}$$

- **The Law of Large Numbers:** Let $X_1, X_2, \ldots, X_n$ be independent identically distributed (iid) Bernoulli random variables, $X_i \sim \text{Ber}(p)$. and denote by $\bar{X}_n = (X_1+\ldots+X_n)/n$ the proportion of successes in these n independent trials. Then

$$P(\left|\bar{X}_n - p\right| < \epsilon) \longrightarrow 1 \qquad \text{as } n \to \infty, \quad \text{for any } \epsilon > 0$$

The proof is an immediate consequence of $\text{var}(\bar{X}_n) = p(1-p)/n$ and Chebychev's inequality.

The same applies for $X_1, X_2, \ldots, X_n$ iid $\sim \text{Pois}(\lambda)$. We saw that $X_1+\ldots+X_n \sim \text{Pois}(n\lambda)$ so that $E(\bar{X}_n) = n\lambda/n = \lambda$ and $\text{var}(\bar{X}_n) = n\lambda/n^2 = \lambda/n$ so that

$$P(\left|\bar{X}_n - \lambda\right| < \epsilon) \longrightarrow 1 \qquad \text{as} \quad n \to \infty, \quad \text{for any } \epsilon > 0$$