Great (Conditional) Expectations

- X and Y are jointly discrete random variables
- Recall, conditional expectation of X given Y = y:

$$E[X \mid Y = y] = \sum_x xP(X = x \mid Y = y) = \sum_x x\, p_{X|Y}(x \mid y)$$

- Analogously, jointly continuous random variables:

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

Computing Probabilities by Conditioning

- X = indicator variable for event A: $X = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$
  - E[X] = P(A)
  - Similarly, E[X | Y = y] = P(A | Y = y) for any Y
  - So: E[X] = E_Y[E_X[X | Y]] = E[E[X | Y]] = E[P(A | Y)]
  - In discrete case:

$$E[X] = \sum_y P(A \mid Y = y)P(Y = y) = P(A)$$

    - Also holds analogously in continuous case
  - Generalize, defining indicator variable F_i = (Y = y_i):

$$P(A) = \sum_{i=1}^{n} P(A \mid F_i)P(F_i)$$

    - Called “Law of total probability”

Hiring Software Engineers

- Interviewing *n* software engineer candidates
  - All *n*! orderings equally likely, but only hiring 1 candidate
    - Claim: There is α-to-1 factor difference in productivity between the “best” and “average” software engineer
    - Depending on who you talk to, usually: 10 < α < 100
  - Right after each interview must decide hire/no hire
  - Feedback from interview of candidate *i* is just relative ranking with respect to previous *i* – 1 candidates
    - Strategy: first interview *k* (of *n*) candidates, then hire next candidate better than all of first *k* candidates
    - P_k(best) = probability that best of all *n* candidates is hired
    - X = position of best candidate (1, 2, ..., *n*)

$$P_k(\text{Best}) = \sum_{i=1}^{n} P_k(\text{Best} \mid X = i)P(X = i) = \frac{1}{n}\sum_{i=1}^{n} P_k(\text{Best} \mid X = i)$$

Hiring Software Engineers (cont.)

- Note: $P_k(\text{Best} \mid X = i) = 0 \quad \text{if} \quad i \le k$
- We will select best candidate (in position *i*) if best of first *i* – 1 candidates is among the first *k* interviewed

$$P_k(\text{Best} \mid X = i) = P_k(\text{best of first } i-1 \text{ in first } k \mid X = i) = \tfrac{k}{i-1} \quad \text{if} \quad i > k$$

$$P_k(\text{Best}) = \frac{1}{n}\sum_{i=1}^{n} P_k(\text{Best} \mid X = i) = \frac{1}{n}\sum_{i=k+1}^{n} \frac{k}{i-1}$$

$$\approx \frac{k}{n}\int_{i=k+1}^{n} \frac{1}{i-1}\, di = \frac{k}{n}\ln(i-1)\Big|_{k+1}^{n} = \frac{k}{n}\ln\frac{n-1}{k} \approx \frac{k}{n}\ln\frac{n}{k}$$

  - To maximize, differentiate P_k(Best) with respect to *k*:

$$g(k) = \frac{k}{n}\ln\frac{n}{k} \qquad g'(k) = \frac{1}{n}\ln\frac{n}{k} + \frac{k}{n}\left(\frac{-1}{k}\right) = \frac{1}{n}\ln\frac{n}{k} - \frac{1}{n}$$

  - Set $g'(k) = 0$ and solve for *k*:

$$\frac{1}{n}\ln\frac{n}{k} - \frac{1}{n} = 0 \;\Rightarrow\; \ln\frac{n}{k} = 1 \;\Rightarrow\; \frac{n}{k} = e \;\Rightarrow\; k = \frac{n}{e}$$

  - Interview n/e candidates, then pick best: P_k(Best) ≈ 1/e ≈ 0.368

Moment Generating Functions

- Moment Generating Function (MGF) of a random variable X, where –∞ < *t* < ∞:

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

  - When X is discrete: $M(t) = \sum_x e^{tx} p(x)$
  - When X is continuous: $M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$
- Oh, that’s nice. Um... why should I care?

Bring on the Moments!

- Start with: $M(t) = E[e^{tX}]$
  - Now differentiate *M*(*t*) with respect to *t*, evaluate at *t* = 0

$$M'(t) = \tfrac{d}{dt}M(t) = \tfrac{d}{dt}E[e^{tX}] = E[\tfrac{d}{dt}(e^{tX})] = E[Xe^{tX}]$$

$$M'(0) = E[Xe^{0}] = E[X]$$

  - That’s pretty neat, let’s do it again:

$$M''(t) = \tfrac{d}{dt}M'(t) = \tfrac{d}{dt}E[Xe^{tX}] = E[\tfrac{d}{dt}(Xe^{tX})] = E[X^2e^{tX}]$$

$$M''(0) = E[X^2e^{0}] = E[X^2]$$

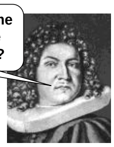
  - Do it as often as you like:

$$M^n(t) = \left(\tfrac{d}{dt}\right)^n M(t) = E[X^n e^{tX}]$$

$$M^n(0) = E[X^n]$$

Let's Take It Out For a Spin

- X ~ Ber(p)



$$M(t) = E[e^{tX}] = \sum_{x=0}^{1} e^{tx} p(x)$$
$$= e^{0}(1-p) + e^{t} p = e^{t} p + 1 - p$$

Sadly, no...

$$M'(t) = e^{t} p \;\Rightarrow\; M'(0) = E[X] = e^{0} p = p$$
$$M''(t) = e^{t} p \;\Rightarrow\; M''(0) = E[X^2] = e^{0} p = p$$
$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1-p)$$

Can You Do That With the Binomial?

- X ~ Bin(n, p)

$$M(t) = E[e^{tX}] = \sum_{k=0}^{n} e^{tk}\binom{n}{k} p^k (1-p)^{n-k}$$
$$= \sum_{k=0}^{n}\binom{n}{k}\left(pe^t\right)^k (1-p)^{n-k} = (pe^t + 1 - p)^n$$

- Binomial theorem: $\sum_{k=0}^{n}\binom{n}{k} x^k y^{n-k} = (x+y)^n \qquad x = pe^t,\ y = (1-p)$

$$M'(t) = n(pe^t + 1 - p)^{n-1} pe^t \Rightarrow M'(0) = E[X] = n(pe^0 + 1 - p)^{n-1} pe^0 = np$$
$$M''(t) = n(n-1)(pe^t + 1 - p)^{n-2}\left(pe^t\right)^2 + n(pe^t + 1 - p)^{n-1} pe^t$$
$$M''(0) = E[X^2] = n(n-1)(1)^{n-2}(p)^2 + n(1)^{n-1} p = n(n-1)p^2 + np$$
$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p)$$

Properties of MGFs

- X and Y are independent random variables
  $$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t)$$
  - Also, if joint MGF factors, then X and Y independent
- MGF uniquely determines distribution
  - Example: $M_X(t) = (0.3e^t + 0.7)^6$
  - Recall MGF for Binomial: $M_X(t) = (pe^t + 1 - p)^n$
  - So: X ~ Bin(6, 0.3)
- Distributions with same MGF are the same!
  $$M_X(t) = M_Y(t) \text{ iff } X \sim Y$$

Joint Moment Generating Functions

- Consider any n random variables $X_1, X_2, \ldots X_n$
  - Joint moment generating function:
    $$M(t_1, t_2, \ldots, t_n) = E[e^{t_1X_1 + t_2X_2 + \ldots + t_nX_n}]$$
  - Individual moment generating functions obtained:
    $$M_{X_i}(t) = E[e^{tX_i}] = M(0, \ldots, 0, t, 0, \ldots, 0) \text{ where } t \text{ at } i\text{th place}$$
  - $X_1, X_2, \ldots X_n$ independent if and only if:
    $$M(t_1, t_2, \ldots, t_n) = M_{X_1}(t_1)M_{X_2}(t_2)\ldots M_{X_n}(t_n)$$
  - Proof:
    $$M(t_1, t_2, \ldots, t_n) = E[e^{t_1X_1 + t_2X_2 + \ldots + t_nX_n}] = E[e^{t_1X_1}e^{t_2X_2}\ldots e^{t_nX_n}]$$
    By independence:
    $$= E[e^{t_1X_1}]E[e^{t_2X_2}]\ldots E[e^{t_nX_n}] = M_{X_1}(t_1)M_{X_2}(t_2)\ldots M_{X_n}(t_n)$$

Generating a Joint Moment

- This is Denise Richards and her family

  - She is Charlie Sheen's ex-wife
  - They generate their moments independently now
    - We can call them independent *random* variables...
  - Yes, I know this slide is gratuitous
    - Sorry... it's the day after the midterm!

Poisson, May I Have a Moment?

- X ~ Poi(λ)

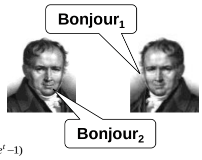
$$M(t) = E[e^{tX}] = \sum_{n=0}^{\infty} e^{tn}\frac{e^{-\lambda}\lambda^n}{n!}$$
$$= e^{-\lambda}\sum_{n=0}^{\infty}\frac{\left(\lambda e^t\right)^n}{n!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

Taylor series: $e^x = \sum_{n=0}^{\infty}\frac{x^n}{n!}$

$$M'(t) = (\lambda e^t)e^{\lambda(e^t-1)} \;\Rightarrow\; M'(0) = E[X] = (\lambda e^0)e^{\lambda(e^0-1)} = \lambda$$
$$M''(t) = (\lambda e^t)^2 e^{\lambda(e^t-1)} + (\lambda e^t)e^{\lambda(e^t-1)} \;\Rightarrow\; M''(0) = E[X^2] = \lambda^2 e^0 + \lambda e^0 = \lambda^2 + \lambda$$
$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

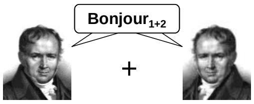
A Tale of Two Poissons

- $X \sim \text{Poi}(\lambda_1)$ $Y \sim \text{Poi}(\lambda_2)$
  - X and Y independent



$$M_{X+Y}(t) = M_X(t) M_Y(t)$$
$$= e^{\lambda_1(e^t - 1)} e^{\lambda_2(e^t - 1)} = e^{(\lambda_1 + \lambda_2)(e^t - 1)}$$

  - So, $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$



MGF of Normal Distribution

- $X \sim N(\mu_1, \sigma_1^2)$

$$M_X(t) = E[e^{tX}] = e^{\left(\frac{\sigma_1^2 t^2}{2} + \mu_1 t\right)}$$

- Yes, it's that important...

You Call That Normal?

- $X \sim N(\mu_1, \sigma_1^2)$ $M_X(t) = E[e^{tX}] = e^{\left(\frac{\sigma_1^2 t^2}{2} + \mu_1 t\right)}$

$$M'_X(t) = (\mu_1 + t\sigma_1^2) e^{\left(\frac{\sigma_1^2 t^2}{2} + \mu_1 t\right)} \Rightarrow M'_X(0) = E[X] = \mu_1$$

$$M''_X(t) = (\mu_1 + t\sigma_1^2)^2 e^{\left(\frac{\sigma_1^2 t^2}{2} + \mu_1 t\right)} + \sigma_1^2 e^{\left(\frac{\sigma_1^2 t^2}{2} + \mu_1 t\right)} \Rightarrow M''_X(0) = E[X^2] = \mu_1^2 + \sigma_1^2$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \mu_1^2 + \sigma_1^2 - \mu_1^2 = \sigma_1^2$$

- Now, $Y \sim N(\mu_2, \sigma_2^2)$ where X and Y independent

$$M_Y(t) = E[e^{tY}] = e^{\left(\frac{\sigma_2^2 t^2}{2} + \mu_2 t\right)}$$

$$M_X(t) M_Y(t) = e^{\left(\frac{\sigma_1^2 t^2}{2} + \mu_1 t\right)} e^{\left(\frac{\sigma_2^2 t^2}{2} + \mu_2 t\right)} = e^{\left(\frac{(\sigma_1^2 + \sigma_2^2) t^2}{2} + (\mu_1 + \mu_2) t\right)} = M_{X+Y}(t)$$

  - Uniquely determines: $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

A Little Surprise Just For You

- X and Y are independent Normal random variables
  - $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\mu, \sigma^2)$
  - Consider random variables: $V = X + Y$ and $W = X - Y$
    - Are V and W independent?
  - Joint MGF of V and W:

$$M(t_1, t_2) = E[e^{t_1 V} e^{t_2 W}] = E[e^{t_1(X+Y)} e^{t_2(X-Y)}] = E[e^{(t_1+t_2)X} e^{(t_1-t_2)Y}]$$
$$= E[e^{(t_1+t_2)X}] E[e^{(t_1-t_2)Y}] \quad \text{since X and Y independent}$$
$$= e^{\mu(t_1+t_2) + \sigma^2(t_1+t_2)^2/2} e^{\mu(t_1-t_2) + \sigma^2(t_1-t_2)^2/2} = e^{2\mu t_1 + 2\sigma^2 t_1^2/2} e^{2\sigma^2 t_2^2/2}$$
$$= E[e^{t_1 A}] E[e^{t_2 B}]$$

  - Consider: independent $A \sim N(2\mu, 2\sigma^2)$ and $B \sim N(0, 2\sigma^2)$
  - Note: $V \sim A$ and $W \sim B$, so V and W are independent!