g. If two random variables are independent, then the joint pdf or pmf is product of marginals

   i. For discrete variables, $p_{X,Y}(x,y) = P(\{X=x\} \cap \{Y=y\}) = P(X=x)\,P(Y=y) = p_X(x)p_Y(y)$

   ii. For continuous variables, decompose $P(\{X \in (x-\delta, x+\delta)\} \cap$

h. Rule: if two random variables are independent, then expectation of product is product of expectations.

   i. Proof for discrete case: $\mathrm{E}[XY] = \Sigma_x \Sigma_y \, xy p_{X,Y}(x,y) = \Sigma_x \Sigma_y \, xy p_X(x) p_Y(y) = \Sigma_x \, x p_X(x) \, \Sigma_y \, y p_Y(y) = \left(\Sigma_x \, x p_X(x)\right)\left(\Sigma_y \, y p_Y(y)\right) = \mathrm{E}[X]\mathrm{E}[Y]$

i. Covariance for independent variables is zero

   i. Proof for continuous case: $\mathrm{Cov}\,[X,Y] = \mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])] = \mathrm{E}[X-\mathrm{E}[X]]\mathrm{E}[Y-\mathrm{E}[Y]] = 0 \times 0 = 0.$

j. Note that $\mathrm{Cov}\,[X,X] = \mathsf{Var}\,[X]$.

k. Note that $\mathsf{Var}\,[aX] = a^2\mathsf{Var}\,[X]$ and $\mathrm{Cov}\,[aX, bY] = ab\mathrm{Cov}\,[X,Y]$.

   i. By previous fact, we need only prove this for covariance.

   ii. For discrete variables, $\mathrm{Cov}\,[aX, bY] =$

$\Sigma_x \Sigma_y (ax)(by) p_{X,Y}(x,y) = ab \Sigma_x \Sigma_y xy p_{X,Y}(x,y)$.

l. $|\mathrm{Cov}[X,Y]| \le \sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}$.

i. Formal proof uses Cauchy-Schwartz inequality.

ii. Heuristic proof: covariance is largest when X and Y line up in same direction.

m. How big is a big covariance?

i. Divide by its maximum and and see how close to 1 you get.

ii. Result is called correlation

iii. For $a, b > 0$, then $\rho[]\, aX, bY = \rho[]\, X, Y$.

8. Conditional pmf and pdf:

a. Discrete case: $P(X = x|Y = y) = P((X = x) \cap (Y = y)) / P($ $p_{X,Y}(x,y)/p_Y(y)$.

b. Continuous case: $P(X \le x | Y \in (y - \delta, y + \delta)) = \int_{-\infty}^{x} \int_{y-\delta}^{y+\delta} f_{X,Y}(w,z)\, dw\, dz / \int_{y-\delta}^{y+\delta} f_Y(z)\, dz \approx (2\delta) \int_{-\infty}^{x} f_{X,Y}(w,y)\, dw / (2\delta f_Y(y)) = \int_{-\infty}^{x} f_{X,Y}(w,y)\, dw / f_Y(y)$ and so $f_{X|Y}(x|y) = f_{X,Y}(x,y)/f_Y(y)$.

: 2.6

Q. Note that $\mathrm{E}[X + Y] = \Sigma_x \Sigma_y (x + y) p_{X,Y}(x,y) =$

$\Sigma_x \Sigma_y x p_{X,Y}(x,y) + \Sigma_x \Sigma_y y p_{X,Y}(x,y) = \Sigma_x x \Sigma_y p_{X,Y}(x,y) + \Sigma_y y \Sigma_x p_{X,Y}(x,y) = \Sigma_x x p_X(x) + \Sigma_y y p_Y(y)$

1. By extension, holds expectation of sum is sum of expectations for larger sums as well.
2. Since $\mathrm{E}[aZ] = a\mathrm{E}[Z]$, then expectation of average is average of expectations.
3. If things being averaged all have same expectation, expectation of average is that value as well.

R. If random variables X_j all have expectation μ then Var $[X_1 + \cdots + X_n] = \mathrm{E}[((X_1 - \mu) + \cdots + (X_n - \mu))^2] = \Sigma_j \mathrm{E}[(X_j - \mu)^2] + \Sigma_{i \neq j} \mathrm{E}[(X_i - \mu)(X_j - \mu)]$

1. If X_j are independent than for $i \neq j$, $\mathrm{E}[(X_i - \mu)(X_j - \mu)] = \mathrm{E}[X_i - \mu]\mathrm{E}[X_j - \mu] = 0$
2. If X_j are independent and each with variance σ^2 then Var $[X_1 + \cdots + X_n] = n\sigma^2$
   a. Var $[\bar{X}] = (1/n)^2 n\sigma^2 = \sigma^2/n$. Hence variance gets smaller as n gets larger.