CS174 Formula Sheet John Canny

Independence: Random variables X, Y are *independent* iff for all values u and v,

$$\Pr[X = u, Y = v] = \Pr[X = u]\Pr[Y = v]$$

Expected Value: The expected value $\mathrm{E}[X] = \overline{X}$ of X:

$$\mathrm{E}[X] = \sum_k k\Pr[X = k] \text{ and if } X \text{ is integer-valued: } \mathrm{E}[X] = \sum_{k=1}^{\infty} \Pr[X \geq k]$$

Linearity of Expectation: Does not require independence:

$$\mathrm{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathrm{E}[X_i]$$

Variance and standard deviation: The variance $\mathrm{Var}(X)$ and standard deviation σ_X are defined as:

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

Covariance: The covariance of two random variables $\mathrm{Cov}(X, Y)$ is defined:

$$\mathrm{Cov}(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$$

and if X, Y are independent, $\mathrm{Cov}(X, Y) = 0$.

Useful Sums:

$$\sum_{k=1}^{n} k^m = \frac{n^{m+1}}{m+1} + O(n^m) \qquad \sum_{k=0}^{\infty} a^k = \frac{1}{1-a} \qquad \sum_{k=1}^{n} \frac{1}{k} = H_n \approx \ln n \qquad \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

Exponentials and Stirling:

$$\left(1 + \frac{1}{n}\right)^n \approx e \qquad \left(1 - \frac{1}{n}\right)^n \approx e^{-1} \qquad n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + O\left(\frac{1}{n^2}\right)\right)$$

Binomial Distribution: with parameters n and p:

$$\Pr[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \text{ and } \binom{n}{k} \text{ is also the coefficient of } x^k \text{ in } (1+x)^n$$

with $\mathrm{E}[X] = np$ and $\mathrm{Var}(x) = np(1-p)$.

Poisson Distribution: with parameter λ:

$$\Pr[X = k] = \frac{e^{-\lambda}\lambda^k}{k!}$$

with $\mathrm{E}[X] = \lambda$ and $\mathrm{Var}(X) = \lambda$.

Geometric Distribution: with parameter p:

$$\Pr[X = k] = (1-p)^{k-1}p$$

where $\mathrm{E}[X] = 1/p$ and $\mathrm{Var}(X) = (1-p)/p^2$.

Markov Bound: For X a non-negative random variable:

$$\Pr[X \geq t] \leq \frac{\mathrm{E}[X]}{t}$$

Chebyshev Bound: For X any random variable:

$$\Pr[|X - \overline{X}| \geq t\sigma_X] \leq \frac{1}{t^2} \qquad \text{or} \qquad \Pr[|X - \overline{X}| \geq s] \leq \frac{\mathrm{Var}(X)}{s^2}$$

Chernoff lower tail bound: For X any random variable which is a sum of independent Poisson trials with $\mathrm{E}[X] = \mu$ and $\delta \in (0,1]$:

$$\Pr[X < (1-\delta)\mu] < \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu} < \exp(-\mu\delta^2/2)$$

Chernoff upper tail bound: For X any random variable which is a sum of independent Poisson trials with $\mathrm{E}[X] = \mu$ and $\delta > 0$:

$$\Pr[X > (1+\delta)\mu] < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} < \begin{cases} \exp(-\mu\delta^2/4) & \text{for } \delta \leq 2e-1 \\ 2^{-\delta\mu} & \text{for } \delta > 2e-1 \end{cases}$$

Inclusion/Exclusion: For a family of events E_i, define $p_i = \Pr[E_i]$, $p_{ij} = \Pr[E_i \wedge E_j]$ and so on. Now define $S_1 = \sum_i p_i$, $S_2 = \sum_{ij} p_{ij}$ etc.

$$\Pr[\text{at least } k \text{ of the } E_i \text{ occur}] = S_k - \binom{k}{k-1}S_{k+1} + \binom{k+1}{k-1}S_{k+2} - \binom{k+2}{k-1}S_{k+3} + \cdots \pm \binom{n-1}{k-1}S_n$$

$$\Pr[\text{exactly } k \text{ of the } E_i \text{ occur}] = S_k - \binom{k+1}{k}S_{k+1} + \binom{k+2}{k}S_{k+2} - \binom{k+3}{k}S_{k+3} + \cdots \pm \binom{n}{k}S_n$$

Birthday paradox: If m balls tossed randomly into n bins, probability of two in one bin is $> 1 - \exp(-m(m-1)/2n)$, and is close to 1 for $m \geq \sqrt{2n}$.

Coupon Collecting: If m balls tossed randomly into n bins, expected number of balls to hit all the bins is $nH_n \approx n \ln n$.

General Occupancy: If m balls tossed into n bins, the number of balls in bin 1 has a binomial distribution with parameters $(m, 1/n)$. For large m, n and small k, it is well approximated by a Poisson distribution with parameter $\lambda = m/n$. In other words, if X is the number of balls in bin 1:

$$\Pr[X = k] = \binom{m}{k}\frac{1}{n}^k\left(1 - \frac{1}{n}\right)^{m-k} \approx \frac{(m/n)^k \exp(-m/n)}{k!}$$