Chebyshev's Theorem

Theorem: The probability that any random variable X will assume a value within k standard deviations of its mean is at least $1 - 1/k^2$. In other words,

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

Notice: a) This is regardless of the specific distribution $f(X)$

b) This is a lower bound not an exact equality.

Proof: $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$

$$= \underbrace{\int_{-\infty}^{\mu - k\sigma} (x-\mu)^2 f(x)\,dx}_{A \geq 0} + \underbrace{\int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)\,dx}_{B \geq 0} + \underbrace{\int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)\,dx}_{C \geq 0}$$

$$\geq A + C$$

Now notice than when $x \leq \mu - k\sigma$ or $x \geq \mu + k\sigma$ we have $|x - \mu| \geq k\sigma \Rightarrow (x-\mu)^2 \geq k^2\sigma^2$

so $\sigma^2 \geq A + C \geq \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(x)\,dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 f(x)\,dx$

$$\sigma^2 \geq k^2\sigma^2 \underbrace{\left(\int_{-\infty}^{\mu-k\sigma} f(x)\,dx + \int_{\mu+k\sigma}^{\infty} f(x)\,dx \right)}_{1 - P(\mu - k\sigma < X < \mu + k\sigma)}$$

$$\frac{1}{k^2} \geq 1 - P(\mu - k\sigma < X < \mu + k\sigma)$$

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - 1/k^2$$

* Chebyshev's theorem is useful when trying to compute a lower limit for a probability when the exact distribution of X is not known but the mean and variance are known.

* This happens in statistics: we make multiple observations of the outcome of a random variable X so we can estimate the mean and variance, but we don't know the exact $f(x)$.

<u>Example</u>: X, $\mu_X = 8$, $\sigma_X^2 = 9$, unknown $f(x)$

center a) $P(-4 < X < 20) = P(\mu_X - 4\sigma_X < X < \mu_X + 4\sigma_X)$ $k=4$

$$\geq 1 - \frac{1}{16}$$

tails b) $P(|X-8| \geq 6) = 1 - \underbrace{P(\mu_X - 2\sigma_X < X < \mu_X + 2\sigma_X)}_{\geq 1 - 1/4}$

so $P(|X-8| \geq 6) \leq 1/4$

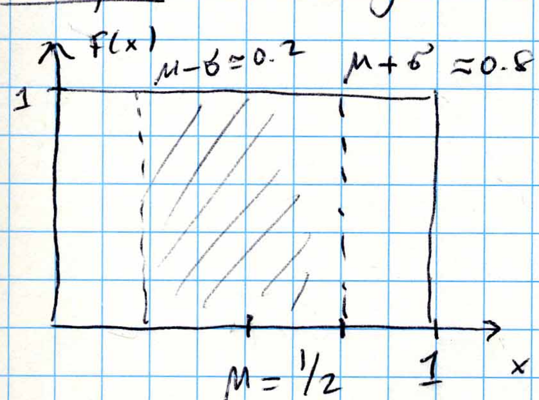
<u>Example</u>: How good is the band for a normal distribution?

From Table A.3 $\quad P(-3 < Z < 3) = 1 - 2 \times 0.0013$
$= 0.9974$

Chebyshev $\quad P(-3 < Z < 3) \geq 1 - \frac{1}{9} \approx 0.89$

Not too bad. (for $f(x)$ with long tails like normal dist)

<u>Example</u>: How good is the band for uniform distribution?



$\sigma^2 = 1/12$ (shown earlier)

$P(\mu - \sigma < X < \mu + \sigma) = 0.6$

Chebyshev:

$P(\mu - \sigma < X < \mu + \sigma) \geq 1 - 1 = 0$

Not useful at all.