## Sets 15 and 16: Section 5.2, The Normal Distribution

**Example Continuous distributions: Checkout duration times in minutes, $X$, have the following cdf:**

$$F(x) = \begin{cases} 0 & x < 0 \\ x^2/4 & 0 \leq x \leq 2 \end{cases}$$

**(a) calculate** $\mathrm{P}(0.5 \leq X \leq 1)$

**(b) calculate the median of** $X$

**(c) calculate the pdf of** $X$

**(d) calculate** $\mathrm{E}(X)$

**The most important distribution in all of Statistics is the normal (Gaussian) distribution.**

**Definition: A rv $X$ has a Normal($\mu, \sigma^2$) distribution if it has pdf**

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2 \right\}$$

**where $x \in \mathscr{R}$, $\mu \in \mathscr{R}$ and $\sigma > 0$.**

**Notes:**

- **the normal is a family of distributions**
- **the density is symmetric about $\mu$**
- **the density never touches zero**
- **the density does not have a closed form integral**
- **the parameters are interpretable:** $\mathrm{E}(X) = \mu$ **and** $\mathrm{Var}(X) = \sigma^2$
- **data are often approximately normal**
- **the standard normal distribution is** $\mathrm{Normal}(0, 1)$ **and is typically represented by the rv** $Z$

**To gain an understanding of the parameters $\mu$ and $\sigma$, sketch plots of the densities:**

- $\mathrm{Normal}(5, 1)$
- $\mathrm{Normal}(7, 1)$
- $\mathrm{Normal}(5, 10)$
- $\mathrm{Normal}(5, 1/10)$

**You must be familiar with the standard normal table (Appendix D, pp 352-3 in text). Calculate:**

**(a)** $\mathrm{P}(Z \leq 3.02)$

**(b)** $\mathrm{P}(Z > 3.03)$

**(c)** $\mathrm{P}(Z < 3.025)$

**(d)** $\mathrm{P}(2.3 \leq Z \leq 2.6)$

**(e)** $\mathrm{P}(Z > -1)$

**(f)** $z^*$ **such that 30.5% of** $Z$**-values exceed** $z^*$

**Proposition: If $X \sim \mathrm{Normal}(\mu, \sigma^2)$, then**

$$Z = \frac{X - \mu}{\sigma} \sim \mathrm{Normal}(0, 1)$$

**Consequence: any normal probability can be converted to a probability for the standard normal. Therefore we only need a single normal table.**

**Example: The number of hours that people watch television is normally distributed with mean 6.0 hours and standard deviation 2.5 hours. What is the probability that a randomly selected person watches more than 8 hours of television per day?**

**Example: The substrate concentration (mg/cm$^3$) of influent to a reactor is normally distributed with $\mu = 0.30$ and $\sigma = 0.06$.**

**(a) What is the probability that the concentration exceeds 0.25?**

**(b) What is the probability that the concentration is at most 0.10?**

**(c) How would you characterize the largest 5% of all concentration values?**

**Proposition: Let $\eta(p)$ denote the $100p$-th percentile of the standard normal distribution. Then the $100p$-th percentile of the $\mathrm{Normal}(\mu, \sigma^2)$ distribution is $\mu + \sigma\eta(p)$.**

**Example: Find the 25.78-th percentile of the $\mathrm{Normal}(5, 100)$.**

**Proposition: Consider $X \sim \mathrm{Bin}(n, p)$ where $np \geq 5$ and $n(1-p) \geq 5$. Then we have the following approximation**

$$X \approx \mathrm{Normal}(np, np(1-p))$$

**The continuity correction below provides a better approximation:**

$$P(X \leq x) \approx P\left(Z \leq \frac{x - np + 0.5}{\sqrt{np(1-p)}}\right)$$

**Example: Obtain** $\mathrm{P}(X \geq 8)$ **where** $X \sim \mathrm{Bin}(10, 1/2)$

**(a) exactly**

**(b) using the normal approximation**

**(c) using the normal approximation with a *continuity correction.***