CSE 312: Foundations of Computing II
Quiz Section #8: Normal Distribution, Central Limit Theorem

Review: Main Theorems and Concepts

Standardizing: Let X be any random variable (discrete or continuous, not necessarily normal), with $\mathbb{E}[X] = \mu$ and $\text{Var}(X) = \sigma^2$. If we let $Y = \frac{X-\mu}{\sigma}$, then $\mathbb{E}[Y] =$ ___________ and $\text{Var}(Y) =$ ___________.

Closure of the Normal Distribution: Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then, $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$. That is, linear transformations of normal random variables are still normal.

"Reproductive" Property of Normals: Let $X_1, \ldots, X_n$ be independent normal random variables with $\mathbb{E}[X_i] = \mu_i$ and $\text{Var}(X_i) = \sigma_i^2$. Let $a_1, \ldots, a_n \in \mathbb{R}$ and $b \in \mathbb{R}$. Then,

$$X = \sum_{i=1}^{n}(a_i X_i + b) \sim \mathcal{N}\left(\sum_{i=1}^{n}(a_i\mu_i + b), \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

There's nothing special about the parameters – the important result here is that the resulting random variable is still normally distributed.

Central Limit Theorem (CLT): Let $X_1, \ldots, X_n$ be iid random variables with $\mathbb{E}[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$. Let $X = \sum_{i=1}^{n} X_i$, which has $\mathbb{E}[X] = n\mu$ and $Var(X) = n\sigma^2$. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, which has $\mathbb{E}\left[\overline{X}\right] = \mu$ and $\text{Var}\left(\overline{X}\right) = \frac{\sigma^2}{n}$. $\overline{X}$ is called the *sample mean*. Then, as $n \to \infty$, $\overline{X}$ approaches the normal distribution $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$. Standardizing, this is equivalent to $Y = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ approaching $\mathcal{N}(0,1)$. Similarly, as $n \to \infty$, X approaches $\mathcal{N}(n\mu, n\sigma^2)$ and $Y' = \frac{X-n\mu}{\sigma\sqrt{n}}$ approaches $\mathcal{N}(0,1)$.

It is no surprise that $\overline{X}$ has mean μ and variance σ^2/n – this can be done with simple calculations. The importance of the CLT is that, for large n, regardless of what distribution X_i comes from, $\overline{X}$ is *approximately normally distributed with mean μ and variance σ^2/n*. Don't forget the continuity correction, only when $X_1, \ldots, X_n$ are discrete random variables.

Markov's Inequality: Let X be a non-negative random variable, and $\alpha > 0$. Then, $\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[X]}{\alpha}$.

Chebyshev's Inequality: Suppose Y is a random variable with $\mathbb{E}[Y] = \mu$ and $\text{Var}(Y) = \sigma^2$. Then, for any $\alpha > 0$, $\mathbb{P}(|Y - \mu| \geq \alpha) \leq \frac{\sigma^2}{\alpha^2}$.

Cantelli's Inequality (one-sided Chebyshev): Suppose Y is a random variable with $\mathbb{E}[Y] = \mu$ and $\text{Var}(Y) = \sigma^2$. Then, for any $\alpha > 0$, $\mathbb{P}(Y - \mu \geq \alpha) \leq \frac{\sigma^2}{\sigma^2+\alpha^2}$.

Chernoff Bound (for the Binomial): Suppose $X \sim Bin(n,p)$ and $\mu = np$. Then, for any $0 < \delta < 1$,

- $\mathbb{P}(X \geq (1+\delta)\mu) \leq e^{-\frac{\delta^2\mu}{3}}$

- $\mathbb{P}(X \le (1-\delta)\mu) \le e^{-\frac{\delta^2 \mu}{2}}$

Exercises

The Φ table is on the last page for use in these exercises.

1. Let $X \sim \mathcal{N}(50, 5)$. What is the probability that X is greater than 45 and less than 52?

2. Before putting any bets down on roulette, you watch 100 rounds, each of which results in an integer between 1 and 36. You count how many rounds have a result that is odd and, if the count exceeds 55, you decide the roulette wheel is unfair. Assuming the roulette wheel is fair, approximate the probability that you make the wrong decision.

3. A factory produces X_i gadgets on day i, where the X_i are independent and identically distributed random variables, each with mean 5 and variance 9.

   (a) Approximate the probability that the total number of gadgets produced in 100 days is less than 440.

   (b) Approximate the greatest value of n such that $P(X_1 + X_2 + \cdots + X_n \ge 5n + 200) \le 0.05$.

4. (a) A fair coin is tossed 50 times. Use the Central Limit Theorem to estimate the probability that fewer than 20 of those tosses come up heads.

   (b) A fair coin is tossed until it comes up heads for the 20th time. Use the Central Limit Theorem to estimate the probability that more than 50 tosses are needed. (Hint: you will need the mean and variance of a geometric random variable, which you can find in Example 2.15 of the text.)

   (c) Compare your answers from parts (a) and (b). Why are they close but not exactly equal?

5. Suppose 59 percent of voters favor Proposition 666. Use the Normal approximation to estimate the probability that a random sample of 100 voters will contain:

   (a) at most 50 in favor.

   (b) between 54 and 64 (inclusive) in favor.

   (c) fewer than 72 in favor.

6. Each day, the probability your computer crashes is 10%, independent of every other day. Approximate the probability of at least 87 crash-free days out of the next 100 days.

7. Suppose $Z = X + Y$, where $X \perp Y$. Z is called the convolution of two random variables. If X, Y, Z are discrete,

$$p_Z(z) = \mathbb{P}(X + Y = z) = \sum_x \mathbb{P}(X = x \cap Y = z - x) = \sum_x p_X(x)\, p_Y(z - x)$$

If X, Y, Z are continuous,

$$F_Z(z) = \mathbb{P}(X + Y \le z) = \int_{-\infty}^{\infty} \mathbb{P}(Y \le z - X \mid X = x) f_X(x) dx = \int_{-\infty}^{\infty} F_Y(z - x) f_X(x) dx$$

Suppose $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

(a) Find an expression for $\mathbb{P}(X_1 < 2X_2)$ using a similar idea to convolution, in terms of $F_{X_1}, F_{X_2}, f_{X_1}, f_{X_2}$. (Your answer will be in the form of a single integral, and requires no calculations – do not evaluate it).

(b) Find s, where $\Phi(s) = \mathbb{P}(X_1 < 2X_2)$ using the "reproductive" property of normal distributions.

8. Suppose $X_1, \ldots, X_n$ are iid $Poi(\lambda)$ random variables, and let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, the sample mean. How large should we choose n to be such that $\mathbb{P}\left(\frac{\lambda}{2} \le \overline{X}_n \le \frac{3\lambda}{2}\right) \ge 0.99$? Use the CLT and give an answer involving $\Phi^{-1}(\cdot)$. Then evaluate it exactly when $\lambda = 1/10$ using the Φ table on the last page.

9. Suppose $X \sim Bin(6, 0.4)$. We will bound $\mathbb{P}(X \ge 4)$ using the tail bounds we've learned, and compare this to the true result.

   (a) Give an upper bound for this probability using Markov's inequality.

   (b) Give an upper bound for this probability using Cantelli's inequality.

   (c) Give an upper bound for this probability using the Chernoff bound.

   (d) Give the exact probability.

10. Let $X \sim \text{Exp}(\lambda)$ and $k > 1/\lambda$.

   (a) Use Markov's inequality to bound $P(X \ge k)$.

   (b) Use Chebyshev's inequality to bound $P(X \ge k)$.

   (c) What is the exact formula for $P(X \ge k)$?

   (d) For $\lambda k \ge 3$, how do the bounds given in parts (a), (b), and (c) compare?

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Table 1: Cumulative distribution function of the standard normal $N(0, 1)$