18.650. Fundamentals of Statistics
Fall 2023. Recitation sheet 1.1

1 Standardization and the Gaussian cdf

Problem 1 Let $X_1, \ldots, X_n$ be i.i.d. with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ and let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Standardize $\bar{X}_n$ (standardizing a random variable Y means finding numbers a, b such that $(Y-a)/b$ has mean 0 and variance 1 (at least approximately)).
Solution: $\mathbb{E}\bar{X}_n = \mathbb{E}\frac{1}{n}\sum_i X_i = \frac{1}{n}\sum_i \mathbb{E}X_i = \mu$ and $\mathbb{V}(\bar{X}_n) = \frac{1}{n^2}\sum_i \mathbb{V}(X_i) = \sigma^2/n$. Then $(\bar{X}_n - \mu)/(\sigma/\sqrt{n})$ has mean 0 and variance 1.

Problem 2 Let $X \sim N(2, 1.96)$. Compute the following probabilities.

1. $\mathbb{P}(X > 1)$.
   Solution: $\mathbb{P}(X > 1) = \mathbb{P}(N(0, 1.4^2) \geq -1) = \mathbb{P}(N(0, 1.4^2) \leq 1) = \Phi(1/1.4) = 0.7611$.
2. $\mathbb{P}(X^3 - 3X^2 + 2X \geq 0)$.
   Solution: factorize as $X(X-1)(X-2)$ to see that it is positive iff $X \in [0,1]$ or $X \geq 2$. Therefore the probability is $0.5 + \Phi(\frac{-1}{1.4}) - \Phi(\frac{-2}{1.4}) = 0.5 + 1 - \Phi(\frac{1}{1.4}) - 1 + \Phi(\frac{2}{1.4}) \approx 0.5 + \Phi(1.42) - \Phi(0.71) \approx 0.5 + 0.9222 - 0.7611 = 0.6611$.
3. Find $a > 1$ such that $\mathbb{P}(1 \leq \frac{|X-2|}{1.4} \leq a) = \alpha$ for some given $\alpha \in (0,1)$.
   Solution: The probability is equal to $\mathbb{P}(N(0,1) \in [-a, -1] \cup [1, a]) = 2(\Phi(a) - \Phi(1))$. Solving for a we get $a = \Phi^{-1}(\Phi(1) + \frac{\alpha}{2})$.

2 Convergence of random variables

Problem 3 (AoS exercise 5.1) Let X_i be i.i.d. with mean μ and variance σ^2. Let

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$$

be the sample variance. Show that $\hat{\sigma}_n^2 \xrightarrow{\mathbb{P}} \sigma^2$. [Hint: write $\hat{\sigma}_n^2$ in terms of $n^{-1}\sum_i X_i^2$ and $\bar{X}_n$.]

Solution: Expanding the square and doing some algebra gives

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\bar{X}_n^2 + \bar{X}_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2.$$

By the LLN, the first average converges in probability to $\mathbb{E}[X_1^2] = \mu^2 + \sigma^2$. By the LLN and the theorem about products of converging sequences, $\bar{X}_n^2$ converges to μ^2. So the difference converges to σ^2.

Problem 4 (AoS exercise 5.9) Suppose that $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$. Define

$$X_n = \begin{cases} X & \text{with probability } 1 - \frac{1}{n}, \\ e^n & \text{with probability } \frac{1}{n} \end{cases} .$$

Does X_n converge to X in probability? In distribution?

Solution: We need to compute $\mathbb{P}(|X_n - X| \geq \epsilon)$ for each $\epsilon > 0$. But the only way $|X_n - X| \geq \epsilon$ is if $X_n \neq X$, which happens with probability $1/n$. Therefore, $\mathbb{P}(|X_n - X| \geq \epsilon) = 1/n \to 0$. So $X_n \xrightarrow{\mathbb{P}} X$ indeed.

It is also true that X_n converges to X in distribution. To see this note that

$$X_n = \begin{cases} -1, & \text{with prob. } \frac{1}{2}(1 - \frac{1}{n}) \\ 1, & \text{with prob. } \frac{1}{2}(1 - \frac{1}{n}), \\ e^n, & \text{with prob. } \frac{1}{n}. \end{cases}$$

This gives rise to the following cdf:

$$\mathbb{P}(X_n \leq x) = \begin{cases} 0, & x < -1 \\ \frac{1}{2}\left(1 - \frac{1}{n}\right), & -1 \leq x < 1, \\ 1 - \frac{1}{n}, & 1 \leq x < e^n, \\ 1, & e^n \leq x. \end{cases}$$

Meanwhile, the cdf of X is

$$\mathbb{P}(X \leq x) = \begin{cases} 0, & x < -1 \\ \frac{1}{2}, & -1 \leq x < 1, \\ 1, & 1 \leq x. \end{cases}$$

Using the above, one can check that for each x, we have $\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$ as $n \to \infty$.

Problem 5 Let $X_1, \dots, X_n$ be $n = 60$ i.i.d. Uniform random variables on the interval $[0, 1]$.

1. Find the parameters μ and σ such that $\bar{X}_n \sim N(\mu, \sigma^2)$ approximately.
   Solution: $\mu = \mathbb{E}\bar{X}_n = \mathbb{E}X_1 = 0.5$ and $\sigma^2 = \mathbb{V}(\bar{X}_n) = \mathbb{V}(X_1)/n = 1/(12n) = 1/720$.
2. What is the approximate distribution of $\exp(\bar{X}_n)$
   Solution: By Thm 5.13 (delta-method) we have $\exp(\bar{X}_n) \approx N(\sqrt{e}, e/720)$.

Problem 6 (AoS exercise 5.13) Let $Z_1, Z_2, \dots$ be positive i.i.d. variables with continuous probability density function f, and suppose $\lambda = \lim_{x \downarrow 0} f(x) > 0$. Let $X_n = n \times \min(Z_1, \dots, Z_n)$. Show that $X_n \rightsquigarrow Exp(1/\lambda)$.

Solution: We will show that for each $x > 0$, the right tail probability $\mathbb{P}(X_n \geq x)$ converges to $e^{-\lambda x}$, which is the right tail probability of $\text{Exp}(1/\lambda)$. This is enough to prove convergence in distribution. Now, we have

$$
\begin{aligned}
\mathbb{P}(X_n \geq x) &= \mathbb{P}(X_1 \geq \frac{x}{n})^n \\
&= (1 - \int_0^{x/n} f(s) ds)^n \\
&= (1 - \int_0^{x/n} [\lambda + ...] \, ds)^n \\
&= (1 - \frac{\lambda x}{n} - ...)^n \to e^{-\lambda x},
\end{aligned}
$$

where the dots represent negligible terms which vanish in the limit.

	Second decimal place of Z									
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

The table lists $P(Z \leq z)$ where $Z \sim N(0,1)$ for positive values of z.