Math-232 — Probability and Statistics

Final exam 2019 —

Exercise 1.

(a) (1 pt) A class of 10 students wants to form a music band with 3 distinct players and each student is willing to learn any instrument. How many different compositions of a band with one guitar, one bass, and one drum can they form?

Solution. Since each student is willing to learn any instrument we have 10 choice for the student who will play guitar, $10 - 1 = 9$ choices for the student who will play bass and $10 - 2 = 8$ choices for the student who will play drum. Then the total number of compositions is

$$10 \times 9 \times 8 = 720.$$

(b) (1 pt) Same question as in **(a)** for a band with two identical guitars, one bass, and one drum.

Solution. A similar reasoning leads easily to the answer

$$\frac{10 \times 9 \times 8 \times 7}{2} = 2520.$$

(c) (1 pt) Let X and Y be i.i.d. uniform random variables on $\{1, 2, ..., 100\}$. What is the probability that $X = Y$?

Solution. We can write this as

$$\begin{aligned}
\mathbb{P}(X = Y) &= \mathbb{P}(\{\omega \in \{1, \dots, 100\} \mid X(\omega) = Y(\omega)\}) \\
&= \mathbb{P}(X = 1 \cap Y = 1) \cup \ldots \cup \mathbb{P}(X = 100 \cap Y = 100) \\
&= \sum_{i=1}^{100} \mathbb{P}(X = i \cap Y = i) \quad (\textit{disjoints events}) \\
&= \sum_{i=1}^{100} \mathbb{P}(X = i)\mathbb{P}(Y = i) \quad (\textit{independance}) \\
&= \frac{1}{100^2} \times 100 \\
&= \frac{1}{100}.
\end{aligned}$$

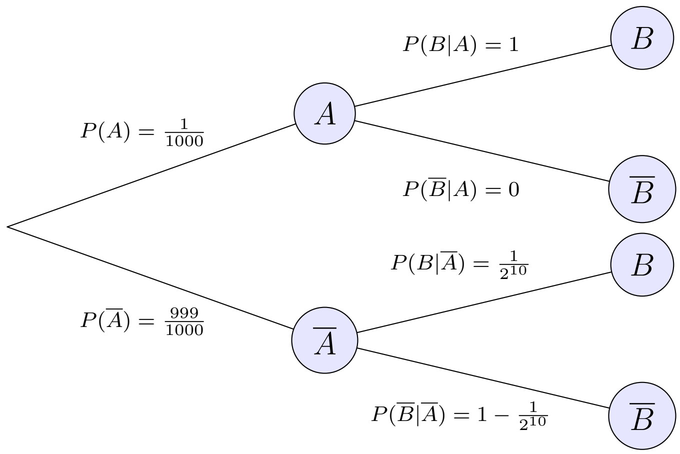
(d) (2 pts) Among 1000 coins, 1 coin has heads on both sides and the other 999 coins are fair (i.e., head or tale with probability 1/2). You pick one coin uniformly at random, and toss it 10 times. Each time, the coin turns up head. What is the probability that the coin you picked is the unfair one?

Solution. Let A and B the events

- A : "We've picked the the fake coin",

- B : "We've made 10 heads in a row".

This situation can be easily represented as



We seek to compute

$$\begin{aligned}
\mathbb{P}(A \mid B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)} \\
&= \frac{1 \times \frac{1}{1000}}{1 \times \frac{1}{1000} + \frac{1}{2^{10}} \times \frac{999}{1000}} \\
&\approx \frac{\frac{1}{2^{10}}}{\frac{1}{2^{10}} + \frac{1}{2^{10}} \times \frac{2^{10}}{2^{10}}} \\
&\approx \frac{1}{2}
\end{aligned}$$

Exercise 2.

(a) (1 pt) Let (X_1, X_2) be a multivariate normal random vector, where each component has mean 0 and variance 1, and $\mathrm{Cov}(X_1, X_2) = \frac{1}{\sqrt{2}}$. Let $W = \sqrt{2}X_1 - X_2$. Are W and X_2 independent ?

Solution. Straightforward calculations gives

$$\begin{aligned}
\mathrm{Cov}(W, X_2) &= \mathrm{Cov}(\sqrt{2}X_1 - X_2, X_2) \\
&= \mathrm{Cov}(\sqrt{2}X_1, X_2) - \mathrm{Cov}(X_2, X_2) \\
&= \sqrt{2}\mathrm{Cov}(X_1, X_2) - \mathrm{Var}(X_2) \\
&= 1 - 1 = 0
\end{aligned}$$

and we conclude that W and X_2 are independent. *Remember that in general we cannot conclude that uncorrelated random variables implies independence, but here since all r.v are (jointly) gaussian it works.*

(b) (2 pts) Let X_1, X_2 be i.i.d. random variables each with mean 0 and variance 1. Let $X_+ = X_1 + X_2$ and $X_- = X_1 - X_2$. Are X_+ and X_- always dependent or always independent? Answer both questions.

Solution. Not always dependent Take $X_1, X_2 \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$. Then we have that

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

hence X_- and X_+ are also multivariate normal. The covariance is

$$\begin{aligned}
\text{Cov}(X_-, X_+) &= \text{Cov}(X_1 + X_2, X_1 - X_2) \\
&= \text{Cov}(X_1, X_1 - X_2) + \text{Cov}(X_2, X_1 - X_2) \\
&= \text{Cov}(X_1 - X_2, X_1) + \text{Cov}(X_1 - X_2, X_2) \\
&= \text{Cov}(X_1, X_1) - \text{Cov}(X_2, X_1) + \text{Cov}(X_1, X_2) - \text{Cov}(X_2, X_2) \\
&= \text{Var}(X_1) - \text{Var}(X_2) \\
&= 1 - 1 = 0
\end{aligned}$$

Not always independent Take $X_1, X_2 \overset{\text{iid}}{\sim} \mathcal{B}(1/2)$. Then the support of X_+ is $\mathcal{S}_1 = \{0, 1, 2\}$ and the support of X_- is $\mathcal{S}_2 = \{-1, 0, 1\}$

$$\begin{cases}
\mathbb{P}(X_+ = 0) = \mathbb{P}(X_1 = 0 \cap X_2 = 0) = \dfrac{1}{4} \\
\mathbb{P}(X_+ = 1) = \mathbb{P}(X_1 = 1 \cap X_2 = 0) \cup \mathbb{P}(X_1 = 0 \cap X_2 = 1) = \dfrac{1}{2} \\
\mathbb{P}(X_+ = 2) = \mathbb{P}(X_1 = 1 \cap X_2 = 1) = \dfrac{1}{4}
\end{cases}$$

and

$$\begin{cases}
\mathbb{P}(X_- = -1) = \mathbb{P}(X_1 = 0 \cap X_2 = 1) = \dfrac{1}{4} \\
\mathbb{P}(X_- = 0) = \mathbb{P}(X_1 = 0 \cap X_2 = 0) \cup \mathbb{P}(X_1 = 1 \cap X_2 = 1) = \dfrac{1}{2} \\
\mathbb{P}(X_- = 1) = \mathbb{P}(X_1 = 1 \cap X_2 = 0) = \dfrac{1}{4}
\end{cases}$$

Let's compute

$$\mathbb{P}(X_- = 1 \cap X_+ = 1) = \mathbb{P}(X_1 = 1 \cap X_2 = 0) = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 0) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$$

. But independence require

$$\mathbb{P}(X_- = 1 \cap X_+ = 1) = \mathbb{P}(X_- = 1)\mathbb{P}(X_+ = 1) = \frac{1}{4}\frac{1}{2} = \frac{1}{8}$$

(c) (1 pt) Can three random variables form a random vector that is multivariate normal such that any two random variables are independent, but the three random variables together are not mutually independent?

Solution. No, if X_1, X_2, X_3 are mutually independent then the vector is multivariate normal such that

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim \mathcal{N}_3 \left(\begin{pmatrix} * \\ * \\ * \end{pmatrix}, \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & * \end{pmatrix} \right)$$

and this variance-covariance matrix implies mutual independence of X_1, X_2 and X_3.

(d) (1 pt) Same question as in **(c)** if the random vector is not required to be multivariate normal.

Solution. Let X_1, X_2 two i.i.d random variable such that $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$. Let $X_3 = X_1 X_2$. We have that (it's a good exercise to check) X_1, X_2 and X_3 are pairwise independent but not mutually independent. We know that $\mathbb{P}(X_3 = 1) = \frac{1}{2}$ so if the r.v are independent the probability $\mathbb{P}(X_1 = 1 \cap X_2 = 1 \cap X_3 = 1)$ must be $\frac{1}{2^3} = 1/8$. In fact we have

$$\begin{aligned}
\mathbb{P}(X_1 = 1 \cap X_2 = 1 \cap X_3 = 1) &= \mathbb{P}(X_1 = 1 \cap X_2 = 1 \cap \{[X_1 = 1 \cap X_2 = 1] \cup [X_1 = -1 \cap X_2 = -1]\}) \\
&= \mathbb{P}(X_1 = 1 \cap X_2 = 1 \cup [(X_1 = 1 \cap X_2 = 1) \cap (X_1 = -1 \cap X_2 = -1)]) \\
&= \mathbb{P}(X_1 = 1 \cap X_2 = 1 \cup \emptyset) \\
&= \mathbb{P}(X_1 = 1 \cap X_2 = 1) \\
&= \frac{1}{2}\frac{1}{2} = \frac{1}{4} \neq \frac{1}{8}.
\end{aligned}$$

Exercise 3. Let $X_1, \ldots, X_n$ be iid random variables such that $\mathbb{P}(X_i = 2) = \mathbb{P}(X_i = -2) = 1/2$ for all $i \in \{1, ..., n\}$.

(a) (2 pts) Prove that $\mathbb{P}(\frac{1}{n}\sum_{i=1}^n X_i > \frac{1}{100})$ tends to 0 as n tends to infinity.

Solution. We see that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i > \frac{1}{100}\right) \leq \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i\right|^2 > \left|\frac{1}{100}\right|^2\right)$$

.
This can be rewrite, by the Markov inequality, as

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i > \frac{1}{100}\right) \leq \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i\right|^2 > \left|\frac{1}{100}\right|^2\right)$$

$$\leq \frac{\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right]}{(1/100)^2}$$

We have that for all $i \in \{1, ..., n\}$

$$\mathbb{E}(X_i) = 0 \qquad \text{and} \qquad \text{Var}(X_i) = 4.$$

Since, the X_i are iid we have that

$$\mathbb{E}(\overline{X}) = 0 \qquad \text{and} \qquad \text{Var}(\overline{X}) = \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] = \frac{\text{Var}(X_1)}{n}$$

This gives

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i > \frac{1}{100}\right) &\leq \frac{\mathbb{E}(\overline{X}^2)}{(1/100)^2} \\ &= \frac{\text{Var}(X_1)}{n(1/100)^2} \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

(b) (1 pt) Consider now the random variable $\tilde{S_n} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$. Compute its moment-generating function, $\mathcal{M}_{\tilde{S_n}} = \mathbb{E}(e^{t\tilde{S_n}})$.

Solution.

$$\begin{aligned}\mathcal{M}_{\tilde{S_n}} = \mathbb{E}(e^{t\tilde{S_n}}) &= \mathbb{E}\left(\exp\left(\frac{t}{\sqrt{n}}\sum_{i=1}^{n} X_i\right)\right) \\ &= \prod_{i=1}^{n} \mathbb{E}\left(\exp\left(\frac{t}{\sqrt{n}} X_i\right)\right) \\ &= \prod_{i=1}^{n}\left(\frac{\exp(2t/\sqrt{n}) + \exp(-2t/\sqrt{n})}{2}\right) \\ &= \left(\frac{\exp(2t/\sqrt{n}) + \exp(-2t/\sqrt{n})}{2}\right)^n.\end{aligned}$$

(c) (2 pts) Using $\mathcal{M}_{\tilde{S_n}}(t)$, find the limiting distribution of $\tilde{S_n}$, i.e., find the distribution of a random variable Z such that $\tilde{S_n} \xrightarrow{D} Z$. Justify your answer.

Hint: For large n, you can use the approximation $\left(\frac{1+\exp\left(\frac{x}{\sqrt{n}}\right)}{2}\right)^n \approx \exp\left(\frac{x\sqrt{n}}{2}+\frac{x^2}{8}\right)$.

Solution.

$$\mathcal{M}_{\tilde{S_n}} = \left(\frac{\exp(2t/\sqrt{n})(1+\exp(-4t/\sqrt{n})}{2}\right)^n = \exp(2nt/\sqrt{n})\left(\frac{1+\exp(-4t/\sqrt{n})}{2}\right)^n$$

For large n we have

$$\begin{aligned}\mathcal{M}_{\tilde{S_n}} &\approx \exp\left(\frac{2nt}{\sqrt{n}}\right)\exp\left(\frac{-4t\sqrt{n}}{2}+\frac{16t^2}{8}\right)\\ &\approx e^{2t^2}\exp\left(\frac{-4tn+4tn}{2\sqrt{n}}\right)\\ &\approx e^{2t^2}\end{aligned}$$

So we have that

$$\mathcal{M}_{\tilde{S_n}} \xrightarrow[n\to\infty]{} e^{2t^2}$$

We recognize the generating function of standard distribution of mean 0 and variance 4. Thus

$$\tilde{S_n} \xrightarrow{\text{d}} \mathcal{N}(0,4)$$

Exercise 4. Let $X_1,\ldots,X_n \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$, with $\theta>0$.

(a) (1 pt) Derive the estimator $\widehat{\theta_1}$ of θ using the method of moments.

Solution.

$$\begin{cases}\mathbb{E}(X_1) &= \theta\\ \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i &= \theta\end{cases} \quad\Longrightarrow\quad \widehat{\theta_1} = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}.$$

(b) (2 pts) Derive the estimator $\widehat{\theta_2}$ of θ using the maximum likelihood method.

Solution. The likelihood is

$$L(\theta) = \prod_{i=1}^n \mathbb{P}(X_i=x_i) = \prod_{i=1}^n \frac{e^{-\theta}\theta^{x_i}}{x_i!} = \frac{e^{-\theta}\theta^{x_1}}{x_1!}\times\ldots\times\frac{e^{-\theta}\theta^{x_n}}{x_n!} = \frac{e^{-n\theta}\theta^{\sum_{i=0}^n x_i}}{\prod_{i=1}^n x_i!}$$

The log-likelihood is therefore

$$\begin{aligned}\log(L(\theta)) = \ell(\theta) &= \log(e^{-n\theta}\theta^{\sum_{i=1}^n x_i}) - \log\left(\prod_{i=1}^n x_i!\right)\\ &= -n\theta + \left(\sum_{i=1}^n x_i\right)\log(\theta) - \log\left(\prod_{i=1}^n x_i!\right)\end{aligned}$$

Derivatives are

$$\frac{\partial \ell}{\partial \theta} = -n + \frac{1}{\theta}\left(\sum_{i=1}^n x_i\right) \quad \text{and} \quad \frac{\partial^2 \ell}{\partial \theta^2} = -\frac{1}{\theta^2}\left(\sum_{i=1}^n x_i\right) < 0$$

Hence,

$$\widehat{\theta_2} = \overline{X}$$

and this a maximum since the second derivatives of ℓ is negative so the log-likelihood is concave.

(c) (2 pts) Compute the bias, the variance, and the mean square error of $\widehat{\theta_2}$. Recall that MSE $= \mathbb{E}((\widehat{\theta_2} - \theta)^2)$

Solution. The bias is

$$b(\widehat{\theta_2}) = \mathbb{E}(\widehat{\theta_2}) - \theta = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) - \theta = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i) - \theta = \theta - \theta = 0.$$

Thus $\widehat{\theta_2}$ is unbiased. Using the independence of the X_i, the variance of $\widehat{\theta_2}$ is

$$\mathrm{Var}(\widehat{\theta_2}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{n}{n^2}\mathrm{Var}(X_1) = \frac{\theta}{n}.$$

Finally we compute the MSE

$$\begin{aligned}\mathrm{MSE} &= \mathbb{E}((\widehat{\theta_2} - \theta)^2)\\ &= \mathbb{E}(\widehat{\theta_2}^2 - 2\widehat{\theta_2}\theta + \theta^2)\\ &= \mathbb{E}(\widehat{\theta_2}^2) + \mathbb{E}(\theta^2) - 2\theta\mathbb{E}(\widehat{\theta_2})\\ &= \mathrm{Var}(\widehat{\theta_2}) + \mathbb{E}(\widehat{\theta_2})^2 + \theta^2 - 2\theta\mathbb{E}(\widehat{\theta_2})\\ &= \mathrm{Var}(\widehat{\theta_2}) + (\mathbb{E}(\widehat{\theta_2}) - \theta)^2\\ &= \mathrm{Var}(\widehat{\theta_2}) + b^2(\widehat{\theta_2})\\ &= \frac{\theta}{n}.\end{aligned}$$

Exercice 5. Let 0^n be the vector of length n that contains only 0's, and let 1^n be the vector of length n that contains only 1's. Assume that n is even.

Consider the following hypothesis testing problem. Under H_0, the random vector Y^n is the output of transmitting 0^n on a binary symmetric channel, i.e., each component of 0^n is flipped independently to a 1 with a probability $p \in (0, 1/2)$ to produce the random vector Y^n. Under H_1, the random vector Y^n is the output of transmitting 1^n on the same channel, i.e., each component of 1^n is flipped independently to a 0 with probability p. Upon observing Y^n, we would like to decide between the hypotheses H_0 and H_1.

Let P_e be the least possible value of the sum of the false positive and false negative error probabilities for this hypothesis testing problem. Recall that the false positive refers to declaring H_1 when the true hypothesis is H_0, and the false negative refers to the reverse case.

(a) (2 pts) Give an optimal test, i.e., how to choose between H_0 and H_1 given Y^n in order to achieve P_e. Justify your answer.

Solution. Let $\mathcal{Y}$ be the set of values of Y^n where we declare H_1. Then

$$P_e = \underbrace{\mathbb{P}(Y^n \in \mathcal{Y} \mid Y^n \sim H_0)}_{\text{false positive}} + \underbrace{\mathbb{P}(Y^n \notin \mathcal{Y} \mid Y^n \sim H_1)}_{\text{false negative}}$$

knowing that

$$\mathcal{Y} = \{Y^n \in \{0,1\}^n \mid P_{H_1}(Y^n) > P_{H_0}(Y^n)\}$$

The likelihood of H_0 and H_1 are

$$\begin{aligned} P_{H_0}(Y^n) &= \prod_{i=1}^{n} p^{Y_i}(1-p)^{1-Y_i} \\ &= p^{Y_1}(1-p)^{1-Y_1} \times \ldots \times p^{Y_n}(1-p)^{1-Y_n} \\ &= p^{\sum_{i=1}^n Y_i}(1-p)^{n-\sum_{i=1}^n Y_i} \\ &= \left(\frac{p}{1-p}\right)^{\sum_{i=1}^n Y_i}(1-p)^n \end{aligned}$$

$$\begin{aligned} P_{H_1}(Y^n) &= \prod_{i=1}^{n} (1-p)^{Y_i} p^{1-Y_i} \\ &= (1-p)^{Y_1} p^{1-Y_1} \times \ldots \times (1-p)^{Y_n} p^{1-Y_n} \\ &= (1-p)^{\sum_{i=1}^n Y_i} p^{n-\sum_{i=1}^n Y_i} \\ &= \left(\frac{1-p}{p}\right)^{\sum_{i=1}^n Y_i} p^n \end{aligned}$$

We have

$$\begin{aligned} &P_{H_0}(Y^n) < P_{H_1}(Y^n) \\ \iff &\left(\frac{p}{1-p}\right)^{\sum_{i=1}^n Y_i}(1-p)^n < \left(\frac{1-p}{p}\right)^{\sum_{i=1}^n Y_i} p^n \\ \iff &\left(\frac{p}{1-p}\right)^{2\sum_{i=1}^n Y_i} < \left(\frac{p}{1-p}\right)^n \end{aligned}$$

This implies that

$$2\sum_{i=1}^{n} Y_i < n \iff \sum_{i=1}^{n} Y_i < \frac{n}{2}$$

Then by the Neyman-Person lemma we can conclude that the optimal test is

$$\begin{cases} \displaystyle\sum_{i=1}^{n} Y_i \leq \frac{n}{2} \Longrightarrow H_0 \\ \displaystyle\sum_{i=1}^{n} Y_i > \frac{n}{2} \Longrightarrow H_1 \end{cases} \iff \text{majority rule}$$

Hence

$$\mathcal{Y} = \left\{ Y^n \in \{0,1\}^n \;\middle|\; \sum_{i=1}^{n} Y_i > \frac{n}{2} \right\}$$

(b) (2 pts) Prove that P_e tends to 0 when n tends to infinity.

Solution.

$$\begin{aligned} P_e &= \mathbb{P}(Y^n \in \mathcal{Y} \mid Y^n \sim H_0) + \mathbb{P}(Y^n \notin \mathcal{Y} \mid Y^n \sim H_1) \\ &= \mathbb{P}\left(\sum_{i=1}^{n} Y_i > \frac{n}{2} \mid Y^n \sim H_0\right) + \mathbb{P}\left(\sum_{i=1}^{n} Y_i \leq \frac{n}{2} \mid Y^n \sim H_1\right) \end{aligned}$$

It's easy to see that the right probability tends to 0. For the left one we have $Y_1, \ldots, Y_n \overset{iid}{\sim} \mathcal{B}(p)$, so

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{n} Y_i > \frac{n}{2}\right) &= \mathbb{P}\left(\sum_{i=1}^{n} Y_i - np > \frac{n}{2} - np\right) \\ &\leq \mathbb{P}\left(\left|\sum_{i=1}^{n} Y_i - np\right| > \left|\frac{n}{2} - np\right|\right) \\ &= \mathbb{P}\left(\left|\sum_{i=1}^{n} Y_i - np\right|^2 > \left(\frac{n}{2} - np\right)^2\right) \\ &\leq \frac{n\mathrm{Var}(X_i)}{(n/2 - np)^2} \xrightarrow[n\to\infty]{} 0. \end{aligned}$$

(c) (1 pt) Does P_e still tend to 0 when n tends to infinity if p depends on n and is given by $p = \frac{1}{2} - \frac{1}{\log(n)}$? Justify your answer.

Solution.

Yes, knowing that the variance of X_i is $p(1-p)$ just plug this and the value for p in the upper bound found in question **(b)** and the new quantity still tends to 0.