Lecture 28: Implicit increasing couplings

Theorem 26.2 of Lecture 26 gave a simple bound on the total variation distance between the distribution of a sum of indicators and an appropriate Poisson distribution, under the assumption of the existence of a coupling with certain monotonicity properties. We gave some examples where such a coupling could be constructed explicitly (isolated vertices and subgraph counts). Here we derive a general condition that implies the existence of such couplings. We first need some terminology.

Definition 28.1. We say a function $f : \{0,1\}^N \to \mathbb{R}$ is increasing (decreasing) if for all $\mathbf{x} = (x_1, \dots, x_N)$ and $\mathbf{y} = (y_1, \dots, y_N)$ such that $x_i \le y_i$ for all $i = 1, \dots, N$, we have $f(\mathbf{x}) \le f(\mathbf{y})$ ($f(\mathbf{x}) \ge f(\mathbf{y})$).

We can now state our main result.

Theorem 28.2. *Let $\mathbf{Y} = (Y_j)_{j=1}^N$ be a finite collection of independent indicators and assume $X_1, \dots, X_m$ are all increasing or all decreasing functions from $\{0,1\}^N$ into $\{0,1\}$. Let $p_i = \mathbb{P}\big(X_i(\mathbf{Y}) = 1\big)$, $W = \sum_{i=1}^m X_i(\mathbf{Y})$ and denote $\mu := \mathbb{E}[W] = \sum_{i=1}^m p_i$. Then*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Po}(\mu)\big) \le \min\{1, \mu^{-1}\}\bigg(\mathrm{Var}(W) - \mu + 2\sum_{i=1}^m p_i^2\bigg).$$

Note that we have already seen some applications where the theorem applies. If $N = \binom{n}{2}$, m is the number of copies of a fixed graph H in the complete graph on n vertices, $\mathbf{Y}$ is the collection of edge indicators for $G \sim \mathrm{ER}(n,p)$, and X_α is the indicator that there is a copy of H at position α in G, then we recover Theorem 26.3. The number of isolated vertices also fits this setup since the indicator that a vertex is isolated is decreasing in the number of edges.

The proof of the theorem follows by showing that there is a coupling satisfying the hypotheses of Theorem 26.2. To prepare for it, we need a couple of lemmas.

Lemma 28.3. *Let $\mathbf{X} = (X_1, \dots, X_m)$ be a vector of Bernoulli random variables, and let $\mathbf{X}^{(i)} = (X_1^{(i)}, \dots, X_m^{(i)})$ be such that $\mathscr{L}(\mathbf{X}^{(i)}) = \mathscr{L}\big(\mathbf{X} | X_i = 1\big)$. Then the following are equivalent:*

1. *There exists a coupling of $\big(\widehat{\mathbf{X}}, \widehat{\mathbf{X}}^{(i)}\big)$ of $\mathscr{L}(\mathbf{X})$ and $\mathscr{L}\big(\mathbf{X}^{(i)}\big)$ such that for all $j = 1, \dots, m$, we have $\widehat{X}_j^{(i)} \ge \widehat{X}_j$.*
2. *For all increasing indicator functions φ, $\mathbb{E}[\varphi(\mathbf{X}^{(i)})] \ge \mathbb{E}[\varphi(\mathbf{X})]$.*
3. *For all increasing indicator functions φ, $\mathrm{Cov}(X_i, \varphi(\mathbf{X})) \ge 0$.*

Proof. We first show 1 and 2 are equivalent. If such a coupling exists, then since φ is increasing, $\varphi(\widehat{\mathbf{X}}^{(i)}) \ge \varphi(\widehat{\mathbf{X}})$ almost surely for this coupling. Thus

$$\mathbb{E}[\varphi(\mathbf{X}^{(i)})] = \mathbb{E}[\varphi(\widehat{\mathbf{X}}^{(i)})] \ge \mathbb{E}[\varphi(\widehat{\mathbf{X}})] = \mathbb{E}[\varphi(\mathbf{X})],$$

as desired. For the other direction, in the case $m = 2$ and $i = 1$ (say), the assumption implies that

$$\mathbb{P}(X_2^{(1)} \ge x) \ge \mathbb{P}(X_2 \ge x),$$

and so $X_2^{(1)} \ge_{\mathrm{st}} X_2$ and the result follows from Proposition 13.3 from Lecture 13, which says that stochastic domination is the same as the existence of monotonic couplings. Recall in the proof of Proposition 13.3, we explicitly constructed the coupling from generalised inverses of CDFs. The general case is more complicated and is usually referred to as Strassen's theorem, and can be proved in our setting using the "max-cut min-flow theorem" as is done in Theorem 4.31 in Chapter 4 of Roch, or by Hall's marriage theorem and approximation (the basic idea is that you have to "marry" the probability mass of $\mathbf{X}^{(i)}$ to that of $\mathbf{X}$ in such a way that the married

masses are always monotonic with respect to the coordinate-wise ordering). We omit this part of the proof because it would take us too far afield.

The equivalence of 2 and 3 easily follows from the calculation:

$$\mathbb{E}[\varphi(\mathbf{X}^{(i)})] = \mathbb{E}[\varphi(\mathbf{X})|X_i = 1] = \mathbb{E}[X_i\varphi(\mathbf{X})|X_i = 1] = \frac{\mathbb{E}[X_i\varphi(\mathbf{X})]}{\mathbb{P}(X_i = 1)}. \qquad \square$$

The next lemma we need is the FKG inequality (adapted to our setting), which says that increasing functions of independent variables are non-negatively correlated.

Lemma 28.4 (FKG inequality)**.** *Let $\mathbf{Y} = (Y_j)_{j=1}^N$ be a finite collection of independent indicators and f and g be both increasing or both decreasing functions on $\{0,1\}^N$. Then*

$$\mathbb{E}[f(\mathbf{Y})g(\mathbf{Y})] \geq \mathbb{E}[f(\mathbf{Y})]\mathbb{E}[g(\mathbf{Y})].$$

Proof. If f and g are decreasing, then $-f$ and $-g$ are increasing, so it is enough to prove the result for f and g both increasing. We use induction on N. For $N = 1$, let Y_1' be an independent copy of Y_1. Since f and g are increasing, we have

$$0 \leq \mathbb{E}\big[\big(f(Y_1) - f(Y_1')\big)\big(g(Y_1) - g(Y_1')\big)\big] = 2\mathbb{E}[f(Y_1)g(Y_1)] - 2\mathbb{E}[f(Y_1)]\mathbb{E}[g(Y_1)],$$

where the inequality is by considering $Y_1 \geq Y_1'$ and $Y_1 < Y_1'$ separately, and the equality is because Y_1, Y_1' are i.i.d. Now assume the inequality holds for any vector of length $k < N$.

First note that for any $y_1, \ldots, y_{N-1} \in \{0,1\}^{N-1}$, the following four functions are increasing:

$$\tilde{f}(x) = f(y_1, \ldots, y_{N-1}, x), \quad \hat{f}(x_1, \ldots, x_{N-1}) = \mathbb{E}[f(x_1, \ldots, x_{N-1}, Y_N)],$$
$$\tilde{g}(x) = g(y_1, \ldots, y_{N-1}, x), \quad \hat{g}(x_1, \ldots, x_{N-1}) = \mathbb{E}[g(x_1, \ldots, x_{N-1}, Y_N)].$$

Thus, using the induction hypothesis for vectors of length one, independence, and that $\tilde{f}$ and $\tilde{g}$ are increasing, we have

$$\mathbb{E}[f(\mathbf{Y})g(\mathbf{Y})|Y_1, \ldots, Y_{N-1}] \geq \mathbb{E}[f(\mathbf{Y})|Y_1, \ldots, Y_{N-1}]\mathbb{E}[g(\mathbf{Y})|Y_1, \ldots, Y_{N-1}].$$

Similarly, now using the induction hypothesis for vectors of length $(N-1)$, we have

$$\begin{aligned}
&\mathbb{E}\big\{\mathbb{E}[f(\mathbf{Y})|Y_1, \ldots, Y_{N-1}]\mathbb{E}[g(\mathbf{Y})|Y_1, \ldots, Y_{N-1}]\big\} \\
&\qquad = \mathbb{E}\{\hat{f}(Y_1, \ldots, Y_{N-1})\hat{g}(Y_1, \ldots, Y_{N-1})\} \\
&\qquad \geq \mathbb{E}[\hat{f}(Y_1, \ldots, Y_{N-1})]\mathbb{E}[\hat{g}(Y_1, \ldots, Y_{N-1})] \\
&\qquad = \mathbb{E}\big\{\mathbb{E}[f(\mathbf{Y})|Y_1, \ldots, Y_{N-1}]\big\}\mathbb{E}\big\{\mathbb{E}[g(\mathbf{Y})|Y_1, \ldots, Y_{N-1}]\big\} \\
&\qquad = \mathbb{E}[f(\mathbf{Y})]\mathbb{E}[g(\mathbf{Y})].
\end{aligned}$$

Using these last two displays, we easily find

$$\begin{aligned}
\mathbb{E}[f(\mathbf{Y})g(\mathbf{Y})] &= \mathbb{E}\big\{\mathbb{E}[f(\mathbf{Y})g(\mathbf{Y})|Y_1, \ldots, Y_{N-1}]\big\} \\
&\geq \mathbb{E}\big\{\mathbb{E}[f(\mathbf{Y})|Y_1, \ldots, Y_{N-1}]\mathbb{E}[g(\mathbf{Y})|Y_1, \ldots, Y_{N-1}]\big\} \\
&\geq \mathbb{E}[f(\mathbf{Y})]\mathbb{E}[g(\mathbf{Y})]. \qquad \square
\end{aligned}$$

We can now prove the main result.

Proof of Theorem 28.2. We show that there exists a size-bias coupling of W satisfying the hypotheses of Theorem 26.2, which implies the result. From Lemma 28.3, it is enough to show that $\mathrm{Cov}(X_i(\mathbf{Y}), \varphi \circ \mathbf{X}(\mathbf{Y})) \geq 0$ for all increasing indicator functions φ. However, since each $X_i(\mathbf{Y})$ is an increasing or decreasing function applied to independent indicators, then so is $\varphi \circ \mathbf{X}(\mathbf{Y})$. Thus we may apply the FKG inequality of Lemma 28.4 with $f = X_i$ and $g = \varphi \circ \mathbf{X}$, which implies

$$\mathbb{E}[X_i(\mathbf{Y})\varphi \circ \mathbf{X}(\mathbf{Y})] \geq \mathbb{E}[X_i(\mathbf{Y})]\mathbb{E}[\varphi \circ \mathbf{X}(\mathbf{Y})],$$

as desired. $\square$

In addition to isolated vertex and subgraph counts in Erdős-Rényi random graphs, the theorem also applies in some new settings where it is not so clear how to construct an explicit monotonic coupling.

Example 28.5 (Large degree vertices)**.** Let d be a non-negative integer and let W be the number of vertices with degree at least d in $G \sim \mathrm{ER}(n,p)$. Clearly W is a sum of indicators that are increasing functions of the edge indicators of the graph G so that Theorem 28.2 can be applied. Letting $q_1 := \sum_{k\geq d} \binom{n-1}{k} p^k (1-p)^{n-k-1}$, we have $\mathbb{E}[W] = nq_1$. Moreover, writing X_i for the indicator that vertex i has degree at least d, we have

$$\mathrm{Var}(W) = nq_1(1-q_1) + n(n-1)\,\mathrm{Cov}(X_1, X_2).$$

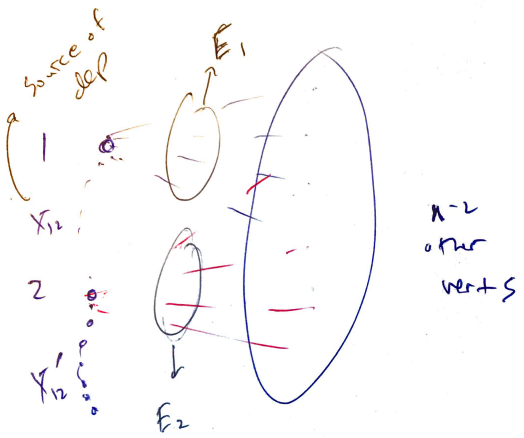
Therefore, applying Theorem 28.2, we find

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Po}(nq_1)\big) \leq q_1 + \frac{(n-1)}{q_1}\big(\mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]\big). \tag{28.1}$$

To compute the covariance, let D_1, D_2 be the degrees of vertices 1 and 2 in G, and we will define $D_2' \stackrel{d}{=} D_2$ on the same space as D_1, D_2, so that D_2' is independent of D_1 and $|D_2 - D_2'| \leq 1$. Then the covariance term is the same as

$$\begin{aligned}&\mathbb{E}\big[\mathbb{I}[D_1 \geq d]\big(\mathbb{I}[D_2 \geq d] - \mathbb{I}[D_2' \geq d]\big)\big] \\ &\qquad = \mathbb{E}\big[\mathbb{I}[D_1 \geq d]\big(\mathbb{I}[D_2 = d, D_2' = d-1] - \mathbb{I}[D_2 = d-1, D_2' = d]\big)\big].\end{aligned} \tag{28.2}$$

The coupling is the one that we used in the proof of Proposition 9.1 in Lecture 9. Copying the notation there, for $i = 1, 2$, let E_i be the number of edges between vertex i and vertices $\{3, \ldots, n\}$, let X_{12} be the indicator that there is an edge between vertices 1 and 2, and let $X_{12}' \sim \mathrm{Be}(p)$ be independent of G. Then for $i = 1, 2$, $D_i = E_i + X_{12}$ and we set $D_2' = E_2 + X_{12}'$. Then note that $D_2' \stackrel{d}{=} D_2$ is independent of D_1 and $|D_2 - D_2'| \leq 1$; see the picture (the same as in Lecture 9).



In this notation, we can write (28.2) as

$$\begin{aligned}&\mathbb{E}\big[\mathbb{I}[D_1 \geq d]\big(\mathbb{I}[D_2 = d, D_2' = d-1] - \mathbb{I}[D_2 = d-1, D_2' = d]\big)\big] \\ &\quad = \mathbb{E}\big[\mathbb{I}[D_1 \geq d]\big(\mathbb{I}[E_2 = d-1, X_{12} = 1, X_{12}' = 0] - \mathbb{I}[E_2 = d-1, X_{12} = 0, X_{12}' = 1]\big)\big] \\ &\quad = \mathbb{P}(E_1 \geq d-1, E_2 = d-1, X_{12} = 1, X_{12}' = 0]) \\ &\qquad\qquad - \mathbb{P}(E_1 \geq d, E_2 = d-1, X_{12} = 0, X_{12}' = 1) \\ &\quad = \mathbb{P}(E_1 \geq d-1)\mathbb{P}(E_2 = d-1)p(1-p) - \mathbb{P}(E_1 \geq d)\mathbb{P}(E_2 = d-1)p(1-p) \\ &\quad = p(1-p)\mathbb{P}(E_1 = d-1)\mathbb{P}(E_2 = d-1)\end{aligned}$$

$$= p(1-p)\left[\binom{n-2}{d-1}p^{d-1}(1-p)^{n-d-1}\right]^2,$$

where we have used the independence of $E_1, E_2, X_{12}, X'_{12}$. Plugging this in to (28.2) and (28.1), we find

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Po}(nq_1)\big) \le q_1 + \frac{d^2(1-p)\left[\binom{n-1}{d}p^d(1-p)^{n-d-1}\right]^2}{(n-1)pq_1}.$$

Example 28.6 (Small degree vertices)**.** Let $d \ge 0$ and let W be the number of vertices with degree at most d in $G \sim \mathrm{ER}(n,p)$. Clearly W is a sum of indicators that are decreasing functions of the edge indicators of the graph G so that Theorem 28.2 can be applied. Moreover, W has the same distribution as the number of vertices having degree at least $(n-d-1)$ in $G' \sim \mathrm{ER}(n,(1-p))$. Thus the previous example with $p \mapsto (1-p)$ and $d \mapsto (n-d-1)$ implies that for $q_2 := \sum_{k\le d}\binom{n-1}{k}p^k(1-p)^{n-k-1}$, we have

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Po}(nq_2)\big) \le q_2 + \frac{(n-d-1)^2p\left[\binom{n-1}{d}p^d(1-p)^{n-d-1}\right]^2}{(n-1)(1-p)q_2}.$$